

C.R.F.E

Access DB# 89382

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: M.A. WALICKA Examiner #: 78201 Date: March 19, 03
Art Unit: 1652 Phone Number 305-7270 Serial Number: 041833, 328
Mail Box and Bldg/Room Location: 10 D06 Results Format Preferred (circle): PAPER DISK E-MAIL
10 D01

If more than one search is submitted, please prioritize searches in order of need. MEJ

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Composition exhibiting a von Willebrand factor etc

Inventors (please provide full names): Saemle et al.

Earliest Priority Filing Date: 11/22/0000

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search SEQ ID NO: 1
SEQ ID NO: 15

and a fusion protein

SEQ ID NO: 1 + SEQ ID NO: 15

N → C

AA Thank you in advance,

1-12
15-136

WALICKA

POINT OF CONTACT:
PAUL SCHULWITZ
TECHNICAL INFO. SPECIALIST
CM1 6806 TEL. (703) 305-1954

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: _____	NA Sequence (#) _____	STN _____
Searcher Phone #: _____	AA Sequence (#) <u>3</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>3/20</u>	Bibliographic _____	Dr.Link _____
Date Completed: <u>3/20</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: <u>10</u>	Fulltext _____	Sequence Systems _____
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: <u>10</u>	Other _____	Other (specify) _____

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:35:07 ; Search time 32 Seconds
(without alignments)
952.968 Million cell updates/sec

Title: SEQID_1_15FUSED
Perfect score: 773
Sequence: 1 AAGGILHLLVAVGPDPVQ.....SPTWSCLITDGFGLGVTI 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 21:

- 1: sp_archaea:
- 2: sp_bacteria:
- 3: sp_fungi:
- 4: sp_human:
- 5: sp_invertebrate:
- 6: sp_mammal:
- 7: sp_mhc:
- 8: sp_organelle:
- 9: sp_phage:
- 10: sp_plant:
- 11: sp_rodent:
- 12: sp_virus:
- 13: sp_vertebrate:
- 14: sp_unclassified:
- 15: sp_rviro:
- 16: sp_bacteriap:
- 17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	773	100.0	1427	4 Q96L37	Q96L37 homo sapien
2	286	37.0	34350	4 Q8WZ42	Q8WZ42 homo sapien
3	272	35.2	26926	4 Q10466	Q10466 homo sapien
4	272	35.2	26926	4 Q8WZB3	Q8WZB3 homo sapien
5	264	34.2	9507	2 Q9EWA1	Q9EWA1 streptomyc
6	264	34.2	15281	3 Q09164	Q09164 tolypocladi
7	243	31.4	10917	2 Q93NW6	Q93NW6 streptomyc
8	241	31.2	9510	2 Q93NX9	Q93NX9 streptomyc
9	241	31.2	16215	5 Q9NFS3	Q9NFS3 drosophila
10	239	30.9	6889	16 Q8XS40	Q8XS40 raietonia s
11	238	30.8	10223	2 Q54296	Q54296 streptomyc
12	237	30.7	9376	2 Q95168	Q95168 pseudomonas
13	236	30.5	11096	2 Q9L4W3	Q9L4W3 streptomyc
14	233	30.1	7829	5 Q18559	Q18559 caenorhabdi
15	231	29.9	9477	2 Q9L4X3	Q9L4X3 streptomyc
16	231	29.9	17352	5 Q95YM2	Q95YM2 procambarus

17	229	29.6	2165	5 Q19791	Q19791 caenorhabdi
18	229	29.6	6396	2 Q9KID7	Q9KID7 streptomyc
19	229	29.6	7463	16 Q924X6	Q924X6 streptomyc
20	226	29.2	5532	2 Q9S0R4	Q9S0R4 streptomyc
21	225	29.1	6048	2 Q93H87	Q93H87 streptomyc
22	225	29.1	7160	5 Q23551	Q23551 caenorhabdi
23	224	29.0	4498	2 Q93HE5	Q93HE5 streptomyc
24	224	29.0	6685	12 Q91W06	Q91W06 transmissib
25	224	29.0	7107	5 Q9VAF7	Q9VAF7 drosophila
26	224	29.0	7576	2 Q9ZGA4	Q9ZGA4 streptomyc
27	223	28.8	1054	5 Q9W493	Q9W493 drosophila
28	222	28.7	6298	11 Q8VHN7	Q8VHN7 mus musculu
29	222	28.7	6781	12 Q91AV2	Q91AV2 porcine epi
30	221	28.6	6831	5 Q23550	Q23550 caenorhabdi
31	220	28.5	6146	2 Q93HJ5	Q93HJ5 streptomyc
32	220	28.5	6420	2 P95814	P95814 streptomyc
33	220	28.5	6875	6 Q28733	Q28733 cryotolegus
34	220	28.5	8563	2 Q54297	Q54297 streptomyc
35	220	28.5	8817	2 Q53840	Q53840 polyangium
36	219	28.3	13055	5 Q09165	Q09165 caenorhabdi
37	218	28.2	5644	2 Q93NK8	Q93NK8 streptomyc
38	217	28.1	6307	4 Q8WKG9	Q8WKG9 homo sapien
39	217	28.1	8243	5 Q96554	Q96554 cryptospori
40	216	27.9	4845	11 Q88738	Q88738 mus musculu
41	216	27.9	5588	2 Q9ALM2	Q9ALM2 saccharopol
42	215	27.8	4349	4 Q9NYQ8	Q9NYQ8 homo sapien
43	215	27.8	4834	4 Q95714	Q95714 homo sapien
44	215	27.8	4898	2 Q93N87	Q93N87 streptomyc
45	214	27.7	4589	11 Q9WU10	Q9WU10 rattus norv

ALIGNMENTS

RESULT 1

Q96L37 PRELIMINARY; PRT; 1427 AA.

ID Q96L37 AC Q96L37; AC Q96L37; PRT; 1427 AA.
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Von Willebrand factor-cleaving protease precursor.
GN ADAMTS13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX PubMed=11557746;
RA Zheng X., Chung D., Takayama T.K., Majerus E.M., Sadler J.E., Fujikawa K.;
RT "Structure of von Willebrand Factor-cleaving Protease (ADAMTS13), a Metalloprotease Involved in Thrombotic Thrombocytopenic Purpura.";
RL J. Biol. Chem. 276:41059-41063(2001).
DR EMBL; AY055376; AAL17652.1; -;
DR MEROPS; M12.241; -;
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn.MTpeptdse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tse_1; 4.
DR PROSITE; PS00215; ADAM_MPRO; 1.
DR PROSITE; PS00092; TSPI; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW SIGNAL 1 33 POTENTIAL.
FT CHAIN 75 1427 VON WILLEBRAND FACTOR-CLEAVING PROTEASE.
SQ SEQUENCE 1427 AA; 153632 MW; EB1BC3AABAC1A4442 CRC64;

Query Match 100.0%; Score 773; DB 4; Length 1427;
Best Local Similarity 100.0%; Pred. No. 1.4e-29;

Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGILHLELLVAVGPDVQFQAQHQDTERYVLTNLNIGAEILLRDPGLGAQFRVHLVKWVIL 60
 Db 75 AAGGILHLELLVAVGPDVQFQAQHQDTERYVLTNLNIGAEILLRDPGLGAQFRVHLVKWVIL 134
 Qy 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRG 120
 Db 135 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRG 194
 Qy 121 VTOLGGACPTWSCLITDTCFDLGVTI 148
 Db 195 VTOLGGACPTWSCLITDTCFDLGVTI 222

RESULT 2

Q8WZ42 PRELIMINARY; PRT: 34350 AA.

AC Q8WZ42;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Titin.
 GN TTN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20309627; PubMed=10850961;
 RA Freiburg A., Trombitas K., Hell W., Cazorla O., Fougereusse F.,
 RA Centner T., Kolmerer B., Witt C., Beckmann J.S., Gregorio C.C.,
 RA Granzier H., Labeit S.;
 RT "Series of exon-skipping events in the elastic spring region of titin
 as the structural basis for myofibrillar elastic diversity.";
 RL Circ. Res. 86:1114-1121(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21573839; PubMed=11717165;
 RA Bang M.L., Centner T., Fornoff F., Geach A.J., Gotthardt M.,
 RA McNabb M., Witt C.C., Labeit D., Gregorio C.C., Granzier H.,
 RA Labeit S.;
 RT "The complete gene sequence of titin, expression of an unusual -700
 kDa titin isoform and its interaction with obscurin identify a novel
 Z-line to I-band linking system.";
 RL Circ. Res. 89:1065-1072(2001).
 DR EMBL; AJ277892; CAD12456.1; .
 DR InterPro; IPR000282; Cytok_receptor_2.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR000577; FGGY_kin.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR001092; HLH_basic.
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003598; Ig_c2.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003596; Ig_v.
 DR InterPro; IPR002016; Peroxidase.
 DR InterPro; IPR004168; PPAK_motif.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00041; fn3; 132.
 DR Pfam; PF00047; Ig; 146.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF02818; PPAK; 53.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00060; FN3; 133.
 DR SMART; SM00409; IG; 167.
 DR SMART; SM00408; IGc2; 148.
 DR SMART; SM00406; IGV; 23.
 DR SMART; SM00220; STKc; 1.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00933; FGGY_KINASES_1; UNKNOWN_1.

DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
 DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
 DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
 DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; UNKNOWN_1.
 SQ SEQUENCE 34350 AA; 3816262 MW; 5B1120058A7CE58A CRC64;

Query Match 37.0%; Score 286; DB 4; Length 34350;
 Best Local Similarity 0.4%; Pred.No.4.1.1;
 Matches 112; Conservative 11; Mismatches 24; Indels 29363; Gaps 32;

Qy 2 AGG----- 4
 Db 3131 AGGNVSTAKLFVEGRDVRIRSIKKEVQVIEKORAVVEFEVNEDDVDAHMKDGEINFGV 3190
 Qy 5 ----- 4

Db 3191 QERHKYVVERRIHRMFISETROSDAGEYTFVAGNRSSVLYVNAPEPOVLOEOPVTV 3250
 Qy 5 ----- 4

Db 3251 QSGKPARCAVISGRPOPKISWYKEQLLSTGFKCKFLHDGOEYTLILLIEAFPEDAAVYT 3310
 Qy 5 ----- 4

Db 3311 CEAKNDYGVAATTSASLSVEVPEVSPQDQMPVYPPIITPLQDTVTSEGPAPFOCRVSG 3370
 Qy 5 ----- 4

Db 3371 TDLKVSWSYKDKKIKPSRFRMTQFEDTVQLEIAEAYPEDEGTYTFVASNAVGVQSSTAN 3430
 Qy 5 -----LELL----- 11
 Db 3431 LSLEAPESILHERIEOEIEWEMKPFSSFLSAEEGLHSAELQLSKINETLELLSPVY 3490
 Qy 12 ----- 11

Db 3491 STKFDSEKGTGPIFIKEVSNADISMGDVATLSVTVIGIPKPIOMFFNGVLLTPSADYK 3550
 Qy 12 ----- 11

Db 3551 FVFDGDDHSLIILFTKLEDEGEYTCASNDYKTCISAYLKINSKGEHGKDTETESAVAK 3610
 Qy 12 ----- 11

Db 3611 SLEKLGPCPPHFLKELKPIRCAOGLPAIFVTVGEPAPTVTWPKENKQLCTSVYYTII 3670
 Qy 12 ----- 11

Db 3671 HNPNGSGTFIVNDPQREDSGLYICKAENMLGESTCAEELLVLLDMDTDPCKAKSTPE 3730
 Qy 12 ----- 11

Db 3731 APEDFPOTPLKGPAAVEALDSEQEIATFVKDTILKAALITEENQQLSVEHIAKANELSSQL 3790
 Qy 12 ----- 11

Db 3791 PLGAQLQSILEQDKLTPESTREFLCINGSIHFOPLKPESPNLQLOIVOSQKTFSEKIL 3850
 Qy 12 ----- 11

Db 3851 MPEEPETQAVLSDETEKIFPSAMSIEQINSULTVEPLKTLAEPSGNYPOSSIEPPMHSYLT 3910
 Qy 12 ----- 11

Db 3911 SVAEVLSPKKTVDNTNREQRTVLQKQEAQSALISLSLAEGHVESLQSPDVMISQVNY 3970
 Qy 12 ----- 11

Db 3971 EPLVPSEHSCTEGKILIESANPLENAGQSAVRIBEGKSLRPFPLALEEKQVLLKEBHD 4030
 Qy 12 ----- 11

Db	4031	NVMPPOOIIESKREPVAIKKQEVQGRDLLSKESLLSGIFEEQRLNLKIQICRALQAAV	4090
Qy	12	-----	11
Db	4091	ASEQGLFSEWLNIEKVEEAVNIQEPHRIMCMYLVTSAKSVTEVTIIIEDVDPQMA	4150
Qy	12	-----	11
Db	4151	NLKMELDALCAIYIEIDILTAEPRIQQGAKTSLQEMDSFGSQKVEPIPEVESK	4210
Qy	12	-----	11
Db	4211	YLISPEEVSFNVQSRVKYLDATPVTKGVASAVUSDEKQDESLEKPEESSESGETEE	4270
Qy	12	-----	11
Db	4271	VATVKIQAEGGFIKEDGPMIHTPLVDTVSEBGDIVHLTTSITNAKEVNVFENKLVPSD	4330
Qy	12	-----	11
Db	4331	EKFCLQDQNTYTLVIDKVNTEHOGYVCEALNDSGKTATSAKLTVVKRAAPVIKRIE	4390
Qy	12	-----	15
Db	4391	PLEVALGHLAKFTCEIQSAPNVRQWFKAGREIYESDKCSIRSSKYISSLEILRTQVWDC	4450
Qy	16	-----	15
Db	4451	GEYCKASNEGVSCTATLTVTEAYPTFLSRPKSLTTFVGAAKFICTVTGTPTVIETI	4510
Qy	16	-----	15
Db	4511	WOKDGAALSPSPNWKISDAENKHILELSNLTIQDRGVYCKASKNKFADICQAEIILIDK	4570
Qy	16	-----	15
Db	4571	PHFIKELEPVQSAINKKHVLECOVDEDRKVTVTWSDGOKLPGDKYKICFEDKIATLEI	4630
Qy	16	-----	15
Db	4631	PLAKLDSGTGYVCTASNEAGSSCSATVTVREPPSFVKVDPSPYLMPOESARLHCKLKG	4690
Qy	16	-----	15
Db	4691	SPVIQVTFWKNKELSESNTVMYFVNSEAILDITDVKVEDSGSYCEAVNDVGSDCST	4750
Qy	16	-----	15
Db	4751	EIVIKEPPSFIKLEPADIVRGTNALLQCEVSGTGPPEISWFKDKKQIRSSKKYRLFQK	4810
Qy	16	-----	15
Db	4811	SLVCLIEFSNADVGEYECVAVNEVGKCGCMATHLLKEPPTFVKVDDLIALGGQTVTL	4870
Qy	16	-----	18
Db	4871	QAAVRGSEPISVTMKQGEVIREDKIKMSFNSGVAVLIIPDVQISFGKYTCIAENEAG	4930
Qy	19	-----	18
Db	4931	SOTSVGELIVKEPAKIIERAELIOVTAGDPATLEVTVAGTPELKPKNYKDGRLVASKKY	4990
Qy	19	-----	18
Db	4991	RISFKNNVAQLKFYSAELHDSQGYTFEISNEVGSSCETTTFTVLDRDIAPFTKPLRND	5050
Qy	19	-----	18
Db	5051	SVVNOTCLDKIAGSLPMRVSWFKDGEIAASDRYIAFVEGTASLEIIRVDMNDAGNF	5110
Qy	19	-----	18
Db	5111	TCRATNSVSKDSSGALIVQEPSPFVTKPGSKDVLPGSAVCLKSTFGOSTPLTIRWFKN	5170
Qy	19	-----	18
Db	5171	KELVSGSGCVITKEALESLELYLVKTSDSGTYTKVSNVAGGVECSANL FVKEPATFVE	5230
Qy	19	-----	18
Db	5231	KLBPSQLKKGDATQACKVTGTPPIKITWFANDREIKESSKHMSFVESTAVLRUTDVG	5290
Qy	19	-----	18
Db	5291	IEDSGEYMCQAONEAGSDHCSSIVIVKESPYTKFKFKPIEVLKEYDVMLLAEVAGTPPFE	5350
Qy	19	-----	18
Db	5351	ITWFKDNTILRSGRKYKFTIQDHLVSLQLILFVAADAGEYOQCRVTNEVGSSICSARVTLR	5410
Qy	19	-----	21
Db	5411	EPSPFIKKIESTSLRGGTAAFOATLKGSLPITVTLKDSDEITEDDNI RMTFENNVA SL	5470
Qy	22	-----	21
Db	5471	YLSGIEVKHDKYVCQAKNDAGIQRCALLSVKEPATITBEAVSIDVTOGDPATLQVKFS	5530
Qy	22	-----	21
Db	5531	GTKEITAKWPKGOELTLGSKYKISVTDVTSILKIISTEKKDSGEYTFEVQNDVGRSCK	5590
Qy	22	-----	21
Db	5591	ARINVLDLIIPSPFTKKLKMDSIKGSFIDLECIVAGSHPISIQWFKDDQEI SASEKYKF	5650
Qy	22	-----	21
Db	5651	SFHDNTAFLEISQLEGDSGTYTCSATNKAGHNOCSGHLTVKEPPYFVKEPQSDVNPT	5710
Qy	22	-----	21
Db	5711	RVQLKALVGGTAPMTIKWFKDNKELHSGAARSVMKDDTSTLSLELFAAKATDSGTYICLS	5770
Qy	22	-----	21
Db	5771	NDVGTATSKATL FVKEPPQFIKKPSVLVLRNGOSTTFECQITGTPKIRVSWYLDGNEIT	5830
Qy	22	-----	21
Db	5831	AIQKHGISFIDGLATFQISGARVENSGTYVCEARN DAGTASCSIELKVKEPPTFIRELKP	5890
Qy	22	-----	21
Db	5891	VEWKYSDVELECEVTGTPPEVTWLKNNREIRSSKKYTLTDRVSVFNLHITKCDPSDTG	5950
Qy	22	-----	21
Db	5951	EYQIVSNEGSCSTRVALKEPPSFIKKIENTTTVLKSSATPOSTVAGSPISITWLK	6010
Qy	22	-----	21
Db	6011	DDQILDEDDNVVISFVDSVATLIQIRSDVNDHSGRYTQCAKNESGVERCYA FLLOEPAQI	6070
Qy	22	-----	21
Db	6071	VEKASVDVTEKDPMTLECVAGTPELKVWLKDGKQI VPSRYFSMSFENNVA SFRIQSV	6130
Qy	22	-----	21
Db	6131	MKODSQYTFKVENDFGSSCCDAYLRLVLDQNI PPSFTKKLT KMDKVLGSSIHMECKVSGS	6190
Qy	22	-----	29
Db	6191	LPISAQWFKDGKEISTSAKYRLVCHERSVLEVNLEEDTANYTCKVSNVAGDDACSGI	6250

QY 30 ----- 29
Db 6251 LTVKEPFLVGRQQAIPDSTVEFKAILKGTTPPKIKWFKDDDELVSQPKCFIGLEGS 6310
QY 30 ----- 29
Db 6311 TSFLNLYSDASKTGQYCHVTNDVSDSCTMLLVTEPPKFVKLEASKIVKAGDSSL 6370
QY 30 ----- 29
Db 6371 ECKIAGSPERVWFRNEHELPAISKYRMTFIDSVAVIQMNUNLSTEDSGDFICEAQNPAG 6430
QY 30 ----- 29
Db 6431 STSCSTKVIKPEPVPFSPPIYETLKNAEVSELELSGTTPPEVWVYKDKQLRSSKKY 6490
QY 30 ----- 29
Db 6491 KIASKNFHTSIHLNVDTSDIGYHCKAQNEVGSDTCVCTVKLKEPPRFVSKLNSLTVA 6550
QY 30 ----- 29
Db 6551 GEPALQASIEGAQPIFVOWLKEKEVIRESENIRITFVENVATLQFAKAPANAGKYIC 6610
QY 30 ----- 29
Db 6611 QIKNDGGMEENMATLMLVLEPAVIVEKAGPMTVTGCTTLECKVAGTPELSVWYKDGKL 6670
QY 30 ----- 29
Db 6671 LTSSQKHKSPYNKISSLRILSVRODAGTYTFOVQNNVGKSCTAVVDVSDRAVPPSFT 6730
QY 30 ----- 29
Db 6731 RRLKNTGGVLCASCILECKVAGSPISVAPHEKTKIVSGAKYQTTFSDNVCTLQLNSLD 6790
QY 30 ----- 32
VLT-----
|||
Db 6791 SSDMGNYTCVAANVAGSDECAVLTVQEPSPFVKPEPELVLPQKNVTFTSVIRGTTPFK 6850
QY 33 ----- 32
Db 6851 VNWFRGARELVKGRCNIFYEDTVAELFNIDISQSEYTCVVSNAGQASCTTRLFVK 6910
QY 33 ----- 32
Db 6911 EPPAFLKRLSDHVSPEGKSIILESTYTGTLPISTWKKDGFNITTEKCNIVTTEKTCIL 6970
QY 33 ----- 32
Db 6971 EILNSTKRDAGQYSCIEAGRDVCGALVSTLEPPYFVTELEPLEAAVGDVSLQCOVA 7030
QY 33 ----- 39
NLNIGAE-----
|||
Db 7031 GTPETVSWYKGDTKLRPTPEYRTFTNNVATLVFNKYNINDSGEYTCKAENSIGTASSK 7090
QY 40 ----- 43
LLRD-----
|||
Db 7091 TVFRIQEROLPSPFARQLKDIETVGLPVTLTCLNGSAPIQVCWYRGDVLRRDENLQT 7150
QY 44 ----- 43
Db 7151 SPVDNVATLKILOTLSHSGQYSCASNPLGTASSARLTAREPKSPFFDIKPVSIDVI 7210
QY 44 ----- 43
Db 7211 AGESADFECHVTGAQPMRITWSKDNKEIRPGGNYTITCVGNTPHLRLKLVGKDSGOYTC 7270
QY 44 ----- 43
Db 7271 QATNDVGKMCQAQLSVKEPPKFVKLEASKVAKOGESIOLECKISGSPFIKVSFRNDS 7330
QY 44 ----- 43

Db 7331 ELHESKWNMSFINSVALLTINEASAEEDSGDYICEAHNGVGDASCSTALTVKAPVFTOK 7390
QY 44 ----- 43
Db 7391 PSPVCAKAGSDVILQCEISGTPPEVWVVKDKOVNRNKKFKITSKHFDTSLHILNLEAS 7450
QY 44 ----- 43
Db 7451 DVGEYHCKATNEVGSDTSCSVKPEPPRFVKLSLSDTSLIGDAVELRAIVEGQFQPSV 7510
QY 44 ----- 43
Db 7511 WLKDRGEVIRESENTRISFIDNIATLQLGSPSEANSKYICQIKNDAGMRECSAVLTVLE 7570
QY 44 ----- 43
Db 7571 PARIIEKPEPMTVTGTPFALECVVGTPELSAKWFKDRELSADSKHHITFINNKVASLK 7630
QY 44 ----- 43
Db 7631 IPCAEMSDKGLYSFEVKNSVGKSNCTSVHVSDDRIVPPSFIRKLKDVNAILGASVVLSCR 7690
QY 44 ----- 43
Db 7691 VSGSAPISVGWFDQGNIEIVSGPKCQSSFSENVCTLNLSLEPSDTGITYTCVAANVAGSDE 7750
QY 44 ----- 43
Db 7751 CSAVLTVQEPSPFQTPDSVEVLPGMSLTFTSVIRGTPPFVKWPKGSRRELVPGESCNIS 7810
QY 44 ----- 43
Db 7811 LEDFVTELELFEVQPLESGDYSLVNDAGSASCTTHLFVKPEPATFVKLADFSVETGSP 7870
QY 44 ----- 43
Db 7871 IVLEATYTGTPISVSWIKDEVLIQSBRCSITMTEKSTILESTIEDYAOVSCLIEN 7930
QY 44 ----- 43
Db 7931 EAQDICEALVLEPPYFIEPLEHVEAVIGEPATLOCKVDGTPETIRISWYKEHTKLRS 7990
QY 44 ----- 43
Db 7991 PAYQOFQKNVASLVINKVDHSDVGEYSCKADNSVAVASSAVLVIKERKLPPFFARKLK 8050
QY 44 ----- 43
Db 8051 DVHETLGPVAPCECRINGSEPLQVSWYKGVLLKDDANLQTSFVHNVTATMILQTDOSHI 8110
QY 44 ----- 43
Db 8111 GOYNCSASNPLGTASSAKLILSEHEVPPFPLKPVSVUDLALGESGTFKCHVTGTAPIKI 8170
QY 44 ----- 43
Db 8171 TWAKDNREIRPGGNYKRWTLVENTATLTVLVKVGKDAGQYTCVASNAGKDCSAHLGVQE 8230
QY 44 ----- 43
Db 8231 PPRFIKLEPSRIVKQDEFTRECKIGSGSPEIKVLWYKDETEIQESSKPRMSPVDSVAVL 8290
QY 44 ----- 43
Db 8291 EMNLSVEDSGDYTCNAHNAAGSASSSTSLKVEPPIFRKKPHPIETLKGADVHLECELO 8350
QY 44 ----- 43
Db 8351 GTPPFHVSWKDKRELRSKKYKINGENFLTSIHLNVDAADIGEYCKATNDVGSDTCV 8410
QY 44 ----- 43

Db	8411	GSIALKAPPRFVKKLSDISTVVGKEVQLQTIEGAEPISVVWFKDKEIVRESNIMISY	8470
Qy	44	-----	43
Db	8471	SENIATLQPSRVEPANAGKYTCQIKNDAGMQECFATLSVLEPATIVEKPESIKVTTGDTG	8530
Qy	44	-----	43
Db	8531	TLECTVAGTPELSTKWFKDQKELTSDNKYKISFFNFKVSGLKLIINVPASDSGVYSEVQNP	8590
Qy	44	-----	43
Db	8591	VGKDSCTASLQVSDRTVPFSFTRKLKETNGLSGSSVMECKVYGSPPISSVWFHEGNEIS	8650
Qy	44	-----	43
Db	8651	SGRKYOTLTNTCALTVMNLEESDSDYTCIATNNAGSDECSAPLTVREPPSFVQKDDP	8710
Qy	44	-----	43
Db	8711	MDVLGTGNTVFTSIVKGTTPFSVSWFKGSELVPGDRCNVSLSDSVAELELFDVDTSSQG	8770
Qy	44	-----	43
Db	8771	EYTCIVSNEAGKASCTHLYIKAPKFKVRLNDYSIEKGKPLILEGTFGTGTPPISVTWK	8830
Qy	44	-----	43
Db	8831	NGINVTQSQRNITTTTEKSAILEIPSTVEDAGQYNCYIENASGKDSQAQILILEPPYF	8890
Qy	44	-----	43
Db	8891	VKOLEPVKVSVDASLQCOLAGTPEIGSVWYKGTDKLRTTTYKMHFRNNVATLVFNQV	8950
Qy	44	-----PSLGAQFR-----	51
Db	8951	DINDSGEYICKAENSVEGSASTFLTVEQKLPFSFSQLRDVQETVGLPVWFDCAISGS	9010
Qy	52	-----	51
Db	9011	EPISVSWYKDGKPLKDSNPVOTSFLDNTATLNFKTRSLAGQVSCATNPIGSSASSAR	9070
Qy	52	-----	51
Db	9071	LILTEGNPPFFDIRLAPVDAVVGESADFECHVTGTQPIKVSMAKDSREIRSGKYQISY	9130
Qy	52	-----	51
Db	9131	LENSAHLTVLKVKGDSQVTCYAVNEVGKDSCTAQLNIKERLIPPSTKRLSETVEETE	9190
Qy	52	-----	51
Db	9191	GNSFKLEGRVAGSQPITVANYKNNIEIQTSPNCEITFNKNTLVLOVRKAGNNDAGLYTCK	9250
Qy	52	-----	51
Db	9251	VSNDAGSALCTSSIVIKEPKPPVFDQHLTPVTVSEGEYVQLSCHVQSGSEPIRQWLKAG	9310
Qy	52	-----	51
Db	9311	REIKPSDRCSFASGTAVLELRDVAKADSGDYVCKASNVAGSDTTKSVTIKDKPAVAP	9370
Qy	52	-----	51
Db	9371	ATKKAADVGRLLFFVSEPOSIRVVEKTTATFTAKVGGDPIPNVKWTKGKWRQLNOGGRVFI	9430
Qy	52	-----	51
Db	9431	HQKGDCAKLEIRDTTKTDGLYRCVAFNEHGEIESNVNLQVDERKKQEKIEGDLRAMLKK	9490
Qy	52	-----	51
Db	9491	TPILKKGAGEEBEIDIMELLKNVDPKEYEKYARMYGITDFRGLLQAFELLKQSQBEETHR	9550
Qy	52	-----	51
Db	9551	LEIEEISERSEDEPEELVSFIQORLSOTEPVTLIKDIENQTVLKNDVAFIDIKINY	9610
Qy	52	-----	51
Db	9611	PEIKLSWYKOTEKLEPSDKFEISIDGDRHRTLVRNQCQKDOGNRYRLVCGPHIASAKLTVI	9670
Qy	52	-----	51
Db	9671	EPAWERHLQDVTLKEGQCTCTMTQCFSPNVKSWFRNGRILKPOGRHKTEVEHKVHLKTI	9730
Qy	52	-----	51
Db	9731	ADVRAEDQGVYCKYEDLETSABLRIEAEPIQFTKRIQNIWSEHQSATPECEVSFDDAI	9790
Qy	52	-----	51
Db	9791	VTWYKGTTELTSOKYNFRNDGRCHYMTIHNVTDDGVSIVARLEPRGEARSTABLYL	9850
Qy	52	-----	51
Db	9851	TTKEIKLELKPPDIPDSRVPITPMPIRAVPPPIPVVAPPILPLLTPEEKKPPPKRIE	9910
Qy	52	-----	51
Db	9911	VTKAVKDAKKVAKPKEMTPREEIVKKPPPTTLIPAKAPEIIDVSSKAEVVKIMTIT	9970
Qy	52	-----	51
Db	9971	RKKEVQKEKEAVYEKKQAVHKEKVFIESPEEPYDELEVEPYTEPPQYVEEDEDYEE	10030
Qy	52	-----	51
Db	10031	IKVEAKKEVHEEWEDEFEQGYEYERBEGYDEGESEWAEYQVEVIOVQKEVYESHER	10090
Qy	52	-----	51
Db	10091	KVPKVPKPKAPPKPKVJIKKPVIEKTSRRMEEEKVQVTKVPEVSKKIVPQKPSRTPV	10150
Qy	52	-----VHLVKMVI-----	59
Db	10151	QEEVIEVKVPVHTKEMVISEEKMFASHTEEEVSVTVPEVQKEIVTEEKIHVAVKRVE	10210
Qy	60	-----	59
Db	10211	PPPVPPELPEKPAPEEAVPPIPKKVPAPKVPKVPKVPKVPKVPKVPKVPKVPKVPK	10270
Qy	60	-----	59
Db	10271	VPEVKKVPPEEKIPVPVAKKEAPPKVPVEQKRVVTEEKITIVTQREESPPPAVPEIP	10330
Qy	60	-----	59
Db	10331	KKKVPEERKVPKSEEVPPPKVPALPKKVPPEEKVAVPVPVAKKAPPPRAEVSKTIV	10390
Qy	60	-----	59
Db	10391	EKKRFVAEEKLSFAVPQORVEVTRHEVSABEWSYSEEEBEGVSI SVYREEEEEEAEVT	10450
Qy	60	-----	59
Db	10451	EYEVMEEPYVVEEKLHIIISKRVEAPAEVTERQEKKIVLKPAPKIEPPPAKVPEA	10510
Qy	60	-----	59
Db	10511	PKKIVPEKKVPAPVKKEKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPK	10570
Qy	60	-----	59
Db	10571	TQBEKVLVAVTKKEAPPKARVPEEPKRAVPEEKVLKLPKREEEPPAKVTEFKRVVKEE	10630

Qy	60	-----	59
Db	10631	KVSTAPKREPOIKVTIMEKERAYTLBEEAVSQREEEYEEYDYKFEVEYPT	10690
Qy	60	-----LTEPE-----	64
Db	10691	EYDQYSEERYERYEHEEYITEPEKPIPVKVPPEPVTKPAPPAKVLKKAPEEK	10750
Qy	65	-----	64
Db	10751	VPVPIPKLKPDPKPPKVEPKKVEKIRISITKKEKEQTEPAKVKPMKPKRVABEKV	10810
Qy	65	-----	64
Db	10811	PVPRKEVAPPVRVPEVPKLEPEPEVAPEEEVVTHTVEEYLVVEEYIHEEEFITEBEV	10870
Qy	65	-----	64
Db	10871	PVIVPKVPEVRPKVPPEEKVPVPVKKEAPPKAPKVPPEPKKVPVLPKKEKPPPAK	10930
Qy	65	-----	64
Db	10931	VPEVPKVPPEKVPVPVKKEAPPKAPKVPPEPKKVPVPAPKKEAPPKAPKEV	10990
Qy	65	-----	64
Db	10991	PKLIPKEKKTPVPKKEAPPKVPKKEPVPVPVVALPQEEVLFEEIYVPEEVLPEE	11050
Qy	65	-----	64
Db	11051	FEVLPEEEVLPPEEEVLPPEEEIPEEEVPPPEEYVPEEFPVPEEVLPEVKPKVPV	11110
Qy	65	-----	64
Db	11111	PAPVPEIKKVTKEKKVTPKKEEAPPKAPKEVPKKEEKRIILPKEEVLPEVTEEPPE	11170
Qy	65	-----	64
Db	11171	EPISBEEIPEBPPSIEEVEAVPRVPEVIKAVPEAPTVPKKEAPPKAVSKKIPEEK	11230
Qy	65	-----	64
Db	11231	VPVQKKEAPPKAPKEVPKKEVPKKEVLPKKEAVPPAKGRTVLEEKVSVAFRQEVVKE	11290
Qy	65	-----	64
Db	11291	RLELEVVEAEVEEIPPEEEFHEVEYFEEGEPHEVEEFIKLEQHRVEBEHRVEKVRVIE	11350
Qy	65	-----	64
Db	11351	VPEAEVEVEFKPAPKPEISEKIIIPKPPPTKVPRKEPPKAPKEVPKKEIIVVEEKVR	11410
Qy	65	-----	64
Db	11411	VPEEPRVPPTKVPDVLPPKEVVPKEVPVPVPAKKEAPPKVPKEAPKEVPKKEVPVPPP	11470
Qy	65	-----	64
Db	11471	KKPEVPPTKVPEVPAKAVEKKVPEAIPPKPESPPEVPPEAPKEVVPPEKVPAPPKKPE	11530
Qy	65	-----	64
Db	11531	VTPVKVPEAPKEVPEKKVPVPPPKPEVPPTKVPEVKVAVPEKKVPEAIPPKPESPPP	11590
Qy	65	-----	64
Db	11591	EVPPEEVEALSEPRAEVVEEPEAPQVTPPKKVPPEKKAPAVAKKPELPPVKVPE	11650
Qy	65	-----	64
Db	11651	VPKEVVPKEKVLVVPKKEAPPKAPKEVPKEVVPPEKVPVPEKVPPEKVPPEVPKKEV	11710
Qy	65	-----	64

Db	11711	LEEKAPVPVPERAESPPEVVEEPEEIAPEEIEEKEKVPVVAEEBEVPPPAVPEEP	11770
Qy	65	-----	64
Db	11771	KKIIPEKKVPVIKKPEAPPKEPEKVEIKPKLPRPPPPPPAPPKEDVKEKIFOLKAI	11830
Qy	65	-----	64
Db	11831	PKKKVPEKQVPEKVELTFLKVPGGGKKVRKLLPERKEPEKEEVVLKSVLRKRPSEEPK	11890
Qy	65	-----	64
Db	11891	VEPKLEKVKKPAVPEPPPPKPEVEVEPTVTKRERKIPEPTKVPEIKPAIPLPAPEPKP	11950
Qy	65	-----CAP-----	67
Db	11951	KPEAEVKTIKPPPEPEPTPIAAPVTVPVVGKKAEPKPEAAKPGPIKGVPKKTPSP	12010
Qy	68	-----	67
Db	12011	IEAERRKLPGSGGKPPDEAPFTYQKAVPLKFVKEIKDIILTESEFVGSSAIFECVLS	12070
Qy	68	-----	67
Db	12071	PSTAITTMMKDSNIRESPKHFRIADGKDKLHIIIDVQLSDAGEYTCVLRGKNEKSTTA	12130
Qy	68	-----	67
Db	12131	KLVEELPVRFVKTLEEEVTVVGOPFLYLSCLNKERDVVRKDGKIIVVEKGRIVPGVI	12190
Qy	68	-----	67
Db	12191	GLMRALTINDADDTAGTIVTVVENANNLECCSVKVEVIRDWLVKPIRDQHVKEPGTA	12250
Qy	68	-----	67
Db	12251	IPACDIADKTPNWKPFYDEIPAEPNDKTEILRDGNHLYLKIKNAMEDIAEYAVEIEG	12310
Qy	68	-----	67
Db	12311	KRYPAKLTLGEREVELLKPIEDVTIYKESASFDABEISEADIPGQWKLKCELLRPSPTCE	12370
Qy	68	-----	67
Db	12371	IKAEGGKRFLTLRKVLDOAGEVLYQALNATITAILTVKEIELDFAVPLKDVTPERRQA	12430
Qy	68	-----	67
Db	12431	RFECVLTREANVIWSKGOIISKSKFDIIADGKHILVINDSQFDDEGVTTAEVEGKKT	12490
Qy	68	-----	67
Db	12491	SARLFTVGTIRLKPMSPLEDOTVKEGETATFVCELSHEKHHVWFKNDAKLHTSRTVLISS	12550
Qy	68	-----	67
Db	12551	EGKTHKLEMEVTLDDISQIAQVKELSSPAQLKVLEADPYFTVKLHDKTAVEKDEITLK	12610
Qy	68	-----	67
Db	12611	CEVSKDVPVKWFKDGBEIVPSPKYSIKADGLRRILKIKADLKDKGEYVCDGTDKTKAN	12670
Qy	68	-----	67
Db	12671	VTVEARLIKVEKPLYGVVEVPGETAHFEIELSEPDVHGQWKLKGOPLTASPCDEIIEGDK	12730
Qy	68	-----	67
Db	12731	KHILILHNCQLGMTGEVVSFOAANAKSAANLKVAKELPLIFITPLSDVKVFEKDEAKFECEV	12790
Qy	68	-----	67

Db 12791 SREPKTFRHLKCTQBITGDDRELIKDGTKGHSWVIKSAFEDAEKMFPEAKDHTSGKLI 12850
Qy 68 ----- 67
Db 12851 IEGIRLKFLTPLKDVTAKEKESAVFTVELSHDNIRKWKFNDRLHTTRSVSMODEGKTH 12910
Qy 68 ----- 67
Db 12911 SITFKDLSIDDSQIRVEAMGMSSEAKLTVLEGDPYFTGKLQDYTGVEKDEVLOCEISK 12970
Qy 68 ----- 67
Db 12971 ADAPVKWFKDGEIKPSKNAVIKADGKKRMLILKALKSDIGQYTCDCGTDKTSGLKDIE 13030
Qy 68 ----- 67
Db 13031 DREIKLVRPLSHVEVNETARFETEISEDDIHANWKLKEALLOTPDCEIKEEGKIHSL 13090
Qy 68 ----- 67
Db 13091 VLHNCRLDOTGGVDFOANVKSANHLRVKPRVIGLLRPLKDVTVTAGETATFDCELSYED 13150
Qy 68 ----- 67
Db 13151 IPVEWYLGCKLEPSDKVVRSEGVHTLTLRDVKLEDAGEVQLTAKDFKTHANLFRKEP 13210
Qy 68 ----- 67
Db 13211 PVEFTKPLEDQVEEGATAVLECEVSRENKVKWFKNGTEILKSKYKYEIVADGRVRKLV 13270
Qy 68 ----- 67
Db 13271 HDCTPEDIKTYTCDAKDFKTSCLNVLVPPHVEFLRPLTDLQVREKEMARFECSELSRENK 13330
Qy 68 ----- 67
Db 13331 VKWFKDGAIEIKKKYDIISKGAVRILVINKCLLDDEAYSCVETARTSGMLTVLEBEA 13390
Qy 68 ----- 67
Db 13391 VFTKNLANIEVSETDTIKLVCEVSKFPAEVIWYKGBEIEIETGRYEILTEGRKRIILVION 13450
Qy 68 ----- 67
Db 13451 AHLEDAGNYNCRPLPSSRTDGVKVKHELAEFISKPNLEILEGEKAEFFVCSIKESFPVQ 13510
Qy 68 ----- 67
Db 13511 WRDDKTLESGDKYDVIADGKRVLVVKDATLQDMGTVMVVGARAAAHLTVIEKLRIV 13570
Qy 68 ----- 67
Db 13571 VPLKDRVKEQEVVFNCEVNTGAKAKWFRNEAIFDSSKYIILOKDLVTLIRDAHL 13630
Qy 68 ----- 67
Db 13631 DDQANNVSLTNHGENVKSAANLIVEEDLRIVEPLKDIETMEKKSVTFWCKVNRNLVT 13690
Qy 68 ----- 67
Db 13691 LKWTKNGEVFPDNRSYRVDKYKMLTIKDCGFPDEGEYIVTAGQDKSVAELLIIIEAPT 13750
Qy 68 ----- 67
Db 13751 EFVSHLEDOTVTEPDADVFCQLSREKANVKWYRNGREIKEGKKYKFEKDGSIHRLI 13810
Qy 68 ----- 67
Db 13811 CRLDDECEYACGVEDRKSRARLFVEEIPVEIIRPPQDILEAPGADVFLAELNKKVEVQ 13870
Qy 68 ----- 67
Db 13871 WLRNNVVVQGDHKHMMSEGIHRLQICDIKPRDQGEYRFTIAKDKEAPAKLELAAPKIK 13930

Qy 68 ----- 67
Db 13931 TADODLVVDVGKPLTMVVPYDAYPKAEAEWFKENEPLSTKTIOTTAQTSFRILEAKGD 13990
Qy 68 ----- 67
Db 13991 KGRYKIVLONKHGKAEGFINLKVIDVPGVRNLEVTEFTFGEVSLAWEPLTDGGSKIIG 14050
Qy 68 ----- 67
Db 14051 YVVERRDIKRTWVLTADRAESCEFTVTGLQGGVEYLFVRSARNRVGTGCEPVETDNVE 14110
Qy 68 ----- 70
Db 14111 ARSKYDVGPPPLNVTITDVNRFGVSLTWEPPEYDGGAEITNYVIELRDKTSIRWDTAMTV 14170
Qy 71 ----- ANLTSS-- 76
Db 14171 RAEDLSATVTDVWEGQYSFRVRAONRIGVCKPSAATPFVKVADPIERPSPVNLTSDDQ 14230
Qy 77 ----- 76
Db 14231 TOSSVOLKWEPPPLKGGSPILGYIIERCEEGKNWIRCNMKLVPELTVKVTGLEKGNKYL 14290
Qy 77 ----- 76
Db 14291 YRVSANENKAGVSDPSEILGPLTADDAFVEPTMDLSAFKDGLEIVPNPITILVPSTGYPR 14350
Qy 77 ----- 76
Db 14351 PTATWCFGDKVLETGDRVMMKTL SAYAELVISPSERSDKGIYTLKLENRVKTIISGEIDVN 14410
Qy 77 ----- 76
Db 14411 VIARPSAPKELKFGDITKDSVHLTWEPDDGGSPLTGYVVEKREVSRTWTKYMDFTVD 14470
Qy 77 ----- LLSVC-- 81
Db 14471 LEFTVPLVQGEYLFVKYCARNKGPGEPAYVDEPVNMSTPATVPDPENVKWRDRTRANS 14530
Qy 82 ----- 81
Db 14531 IFLTWDPPKNDGSRIGKIYIVERCPRGSDKWKVACGEPVAETKMEVTGLEGKWYAYRKA 14590
Qy 82 ----- 81
Db 14591 LNRQASKPSRPTIEIOAVDQEAPEIFLDVKLLAGLTWKAGTKIELPATVTGKPEPKIT 14650
Qy 82 ----- 81
Db 14651 WTKADMILKQDKRITITENVPKSTVTIYDSKRSDTGTVII EAVNVCGRATAVVEVNVLDK 14710
Qy 82 ----- 81
Db 14711 PGPPAAFDITDVNESCCLLTWNPPRDDGSKITNYVERRATDSEVMHKLSTVKDTNFK 14770
Qy 82 ----- 81
Db 14771 ATKLI PNKEYIFRVAENMYGVGPVQASPI TAKYQFDPGPPTRLEPSDITKDAVTLTW 14830
Qy 82 ----- 81
Db 14831 CEPDDGGSPITGYWVERLDPTDKWRCNKMVPKDTTYRVKGLTNKKKYRFRVLAENLA 14890
Qy 82 ----- 81
Db 14891 GPGKSKSTEPILIKDPTDPPWPGKPTVKOVGKTSVRLNMTKPEHDGGAIESYVIELM 14950
Qy 82 ----- 81
Db 14951 KTGTDENVRVAGVPTTQHLPLGLMEGOYSFRVRAVNKAGESEPSDPVLCREKLVP 15010

Qy	82	-----	81
Db	15011	PSPRWLEVINITKNTADLKWTPEKDGSGPITNIYIEXEDVRKGMQTVDTTVKDTKCT	15070
Qy	82	-----	81
Db	15071	VTPLTEGLSVFRAAENAIGOSDYTEIEDSVLAKDTFTTPGPPYALAVVDVTKRHVDLK	15130
Qy	82	-----	81
Db	15131	WEPPKNDGGRPIORYVIEKKERLGRVWKAGKTAGPCNFRVTDVIEGTEVOFVRAENE	15190
Qy	82	-----	81
Db	15191	AGVHPSEPTIELSIEDTPSPPLDLHVTDAGRKHIAIAWKPEKNGGSPIIGYHVM	15250
Qy	82	-----	81
Db	15251	CPVGTEKMRVNSRPIKDLKFVEGVDPKEYVLRVRAVNAIGVSEPSISENVAKDP	15310
Qy	82	-----	81
Db	15311	DKPTIOLETHDIIVIEGEKLSIPVPFRVAVPVTVMHKGKVKASDRLTMKNDHLSAH	15370
Qy	82	-----	81
Db	15371	LEVPKSVRADAGIVTITLENKLGSAASINVKVLPGPCCKDIKASDITKSSCKLTWEP	15430
Qy	82	-----	81
Db	15431	EFDGGTPILVYLERREAGRRTYIPVMGENKLSWTVKDLIPNGEYFFRVKAVNKVGGE	15490
Qy	82	-----	81
Db	15491	YIELKNPVIAQDPKQPPDPVDVEVHNPTAEAMTITWKPLPYDGGSKIMGYIEKIAKE	15550
Qy	82	-----	81
Db	15551	ERWKRNEHLVPILTYTAKGLEEGKEYQFVRAENAAGISEPSRATPPTKAVDPIDAPKV	15610
Qy	82	-----	81
Db	15611	ILRTSVKRGDEIADASISGSPYPTITWIKDENIVPEIEKRAAPLVRRRKGEVQEE	15670
Qy	82	-----	81
Db	15671	EPFVLPLTQRLSDNSKKGESQLRVDRSLRDPDHGLYMIKVENDHGIAPCTVSVLDTPG	15730
Qy	82	-----	81
Db	15731	PPINFEFDIRKTSVLCKWEPLDDGSEIINYTLKKDKTKPDSEMIIVTSTLRHKYS	15790
Qy	82	-----	81
Db	15791	VTKLIEGKEYLFRVAENRFGGPPCVSKPLVAKDPFGPDAPDKPIVEDVTSNMLVKW	15850
Qy	82	-----	81
Db	15851	NEPKDNGSPILGWLEKREVNSTHWSRVNKSLLNALKANVDGLLEGLTYVFRVCAENAAG	15910
Qy	82	-----	81
Db	15911	PGKFSPPDPKTAHDPIGPPGPIPRVTDTSSTTIELEWEPFAPNGGGEIVGYFVDKOLV	15970
Qy	82	-----	81
Db	15971	GTNEWSRCTEKMIKRVQYTVKIEIREGADYKLRVSAVNAAGEPPGPGTPTVAEPQEP	16030
Qy	82	-----	81
Db	16031	VELDVSVKGGIOIMAGKTLRIPAVVTGRPVPTKWTKEGELDKORVIDNVGTSKELI	16090
Qy	82	-----	81

Db	16091	KDALRKHGRVYITATNSCGSKFAAARVEVFDVPGVLDLKPVVNTRKCLLNWSDPEDD	16150
Qy	82	-----	81
Db	16151	GGSEITGFIIERKDAKMTWRQPIETERSKCDITGLLEGOEYKPRVIAKNKFGCGPPVEI	16210
Qy	82	-----	81
Db	16211	GPILAVDPLGPPTSPERLTYTERTKSTITLDWKEPRNGSGSPIQGYIIEKRHKDKPDER	16270
Qy	82	-----	81
Db	16271	VNRLCPTTSFLVENLDEHOMYEPRVKAVNEIGSESEPLPLNVVIQDDEVPTIKRLSV	16330
Qy	82	-----	81
Db	16331	RGDTIKVAGEPVHIPADVTGLPMPKIEWSKNETVIEKPTDALQITKEEVSRSKTELS	16390
Qy	82	-----	81
Db	16391	IPKAVREDKGTVTYTASNRLGSVFRNVHVEVYDRPSPPRNLAVTDIKAESCYLTDAPLD	16450
Qy	82	-----	81
Db	16451	NGGSEITHYVIDKRDAKRKAEEVNTAVEKRYGIWKLIPNGOYEPRVAVNKYGISD	16510
Qy	82	-----	81
Db	16511	ECKSKWVIQDPYRUPGPGPKVLARTKGSMLVSWTPLLNGGSGPITGYWLEKREBGP	16570
Qy	82	-----	81
Db	16571	YMSRVSRAPITKVGLUGVEFNVPRLLLEGVYQFRAMAINAAGIGPPSEPSDEVAGDPF	16630
Qy	82	-----GW-----	83
Db	16631	PGPPSCPEVKDKTKSSISLGMKPPAKDGGSPIKGYIVEMOEBEGTTDKRVNPEDKLITT	16690
Qy	84	-----	83
Db	16691	CECVPNLKLRYKRVKAVNEAGESEPSDTTGEIPATDIOEEPEVFIDIGAQCCLVCK	16750
Qy	84	-----	83
Db	16751	AGSQIRIPAVIKGRPTPKSSWEFDGKAKKMDGVHDIPEDAQLETAENSSVILIIPECKR	16810
Qy	84	-----	83
Db	16811	SHTGKYSITAKNAGQKANTANCRVKVMDVPGPKDLKVSIDITRGCRLSMMPDDGGDRI	16870
Qy	84	-----	83
Db	16871	KGYVIEKRTIDGKAMTKVNPDCGSTTFFVVDLLSEQQYFPRVRAENRFGIGPPVETIQT	16930
Qy	84	-----	83
Db	16931	TARDPIYPPDPPIKLKIGLITKNTVHLSWKPPKNDGSPVTHYIVECLAWDPTGTKEAW	16990
Qy	84	-----	83
Db	16991	RQCNKRDEVELOFTVEDLVEGGEYPRVKAVNAAGVSKPSATVGPDCQCORPMPPSIDLK	17050
Qy	84	-----	83
Db	17051	EFMEVEBEGTNVINIAKIKGVFPFPLTWFKAPPKKPNKEBPVLDTHVKNLVDDTCLVI	17110
Qy	84	-----	83
Db	17111	QSRSDTGLYTTITAVNNLGTASKEMRLNVLGRPGPPVGPIKESVADQMTLSWFPKPD	17170
Qy	84	-----	83

seqid_1_15fused.rspt

Thu Mar 20 13:50:31 2003

Db	17171	DGSKITNYVIEKREANRKTWHVSSPEKCTYITPKLLEGHEYVFRIMAQNKYGIGEP	17230
Qy	84	-----SOTIN-----	88
Db	17231	DSEPETARNLFSVPCAPDKPTVSSVTRNSMTVNWEEPEYDGGSPVTGYWLEMKDITTSKR	17290
Qy	89	-----	88
Db	17291	KRVNRDPIKAMTLGVSQYKVTGLIEGSDYQFRVYAINAAGVGPASLPSPATARDPIAPP	17350
Qy	89	-----	88
Db	17351	PFPKVTWTKSSADLEWSPLKDGSKVTGYIYEKSEGEWEKGDKEVRGKLVVT	17410
Qy	89	-----	88
Db	17411	GLKEGAFYKFRYSAVNIAGIGEPGEVTDVEMKDRLVSPDLQDASVRDRIVVHAGVIR	17470
Qy	89	-----	88
Db	17471	IYVSGKPPPTVTNMNERTLPQETIETTAISSMWIKNCORSHQGVYSLAKNEAGE	17530
Qy	89	-----PEDD-----	92
Db	17531	RKKTIIIVDLVDPGCVGTFFLAHNLNESCULTWFSPEDDGGSPITNYVIEKRESDRRAW	17590
Qy	93	-----	92
Db	17591	TPVTYVTRQNATVQGLIOGKAYFRPRAAENSIGMPFVETSEALVIREPITVPERPED	17650
Qy	93	-----	92
Db	17651	EVKEVTNTVTLTWNPPKYDGGSEIINYVLESRLIGTEKFKVTNDNLLSRKYTVKGLKE	17710
Qy	93	-----	92
Db	17711	GDTYEVRSAVNIVGOKPSFCTKPTCKDELAPPTLHLDKLTIRVGEAFALTGRYS	17770
Qy	93	-----	92
Db	17771	GKPKVSWFKDEADVLEDDRTHIKTTPATLALEKIKAKRSDSGKYCVVVENSTGSRKGF	17830
Qy	93	-----	92
Db	17831	QOVNVDRPGPVPVGSFDEVTQDMVISMKPPLDGGSKITNYIIEKKEVGKDVMPVT	17890
Qy	93	-----	92
Db	17891	SASAKTTCKVSKLEGKDYIFRIHAENLYGISDPLVSDSMKAKDRFRVPDAPDQPIVEV	17950
Qy	93	-----	92
Db	17951	TKDSALVTWKNPHDGGKPTNYILEKRETMSKRWARTKOPHPTKFRVPDILLEGQYE	18010
Qy	93	-----	92
Db	18011	PRVAENEIGDPSPPKPVFAKDIAPKPSPPVPEAIDITTCNSVLTWQPPRHGGSK	18070
Qy	93	-----TDPGH-----	97
Db	18071	ILGYIYEQKVBDEWRANHTPESCPTYKYVTGLRDQTYKPRVLAVNAAGESDPAHV	18130
Qy	98	-----	97
Db	18131	PEPVLVKORLEPPELILDANWAREQHIKVGDTLRLSAILKIGVPPKVTWKEDRDAPKA	18190
Qy	98	-----	97
Db	18191	RIDVTPVGSKLEIRNAHEDGGIYSLTVENPAGSKTVSKVLVDKPGPPRDLEVSEIRK	18250
Qy	98	-----	97
Db	18251	DSCYLTKWEPLDDGGSVITNYVVERRDVASAWSPLSATSKKSHFAKHLNEGNOYLFV	18310
Qy	98	-----	97
Db	18311	AAENQYGRGPFVETPKPIKALDPLHPPGPPKOLHHVDVKTEVSLVMNKPDRDGGSPITG	18370
Qy	98	-----	97
Db	18371	YLVEYOBEGTQDWIKFKFTVTNLECVVTGLOOGKTYFRVKAENIVGLGLPDTTPIECQE	18430
Qy	98	-----	97
Db	18431	KLVPSPVELDKLIEGLVVKAGTTVRFPATIRGVVPVPTAKWTTDGSSEIKTDEHYTVETDN	18490
Qy	98	-----	97
Db	18491	FSSVLTIKCLRRDTGEYQITVSNAGSKTVAVHLTVLDVPGPPTGPIINILDVTPHEMTI	18550
Qy	98	-----	97
Db	18551	SWPPKDDGGSPVINYIVEKODTRKDTGWVVSSGSSKTKLKI PHLQKCEVVFVRVAENK	18610
Qy	98	-----	97
Db	18611	IGVCPILDSTPTVAKHKFSPSPGPKPVWTDITENAAATVSWTLPSDGGSPITGYMYERR	18670
Qy	98	-----	97
Db	18671	EVTGKWRVVKTPIDILKFRVTGLYEGTYEFRVFAENLAGLSPSPSSDPIKACRPIKP	18730
Qy	98	-----ADLV-----	101
Db	18731	PGPPINPKLKDOKSRETADLVWTKPLSDGGSPILGVVVECCQPGTAOMNRINKDELIRQCA	18790
Qy	102	-----	101
Db	18791	FRVPLGIEGNEVRFRIKAANIVGESEPRELAESVIAKDILHPEPELVDVTCRDVITVRVG	18850
Qy	102	-----	101
Db	18851	QTIIRILAVKGRPEPDIWTKEGKVLVREKRVLDIQLPRVELQIKEAVRADHGKYIISA	18910
Qy	102	-----	101
Db	18911	KNSSHAOGSALVNVLDPRPGCONLKVNTVKENTISWENPLDNGSGSEITNFIVEYRKP	18970
Qy	102	-----	101
Db	18971	NQKGSIIVASDVTKRLIKANLLANNEYIFRVCAENKVGVGPTIETKTPILAINPIDRPG	19030
Qy	102	-----	101
Db	19031	PENLHIADKGTFFVYLKWRRPDYDGGSPNLSYHVERRLKGSDDERVHKSGIKETHYMYD	19090
Qy	102	-----	101
Db	19091	RVENQIYEFVQTKNEGESDMVKTEVVVKEDLQKPVLDLKLSGVLTAKAGDTIRLEA	19150
Qy	102	-----	101
Db	19151	GVRGKPFPEVAWTKOKDATDLTRSPRVKIDTRADSKFSLTKAKRSDGKGYVVTATNTAG	19210
Qy	102	-----	101
Db	19211	SFVAYATVNVLDKPGPVRLKIVDVSSDRCTVCMWDPEDDGGCEIQNYILEKCEKRMVM	19270
Qy	102	-----	101
Db	19271	STYSATVLTPTGTTVTRLIEGNEYIFRVRAENKIGTPPTESKPVIAKTKYDKPRDPPE	19330
Qy	102	-----	101
Db	19331	VTKVSKEMTVVMNPPPEYDGGKSIITGYFLEKKEKHSTRVFNKSAIPERRMKVQNLPLD	19390

QY	102	-----	101
Db	19391	HEYQFRVKAENEIGIGEPSLPSRPVAKDPIEPGPTNFRVVDTTKHSITLWGKPVYD	19450
QY	102	-----	101
Db	19451	GGAPIIGYVVMRPKIADASPDEGKRCNAQAOLVRKEFTVTSLENQEYEFVRCANOV	19510
QY	102	-----	101
Db	19511	GICRPALKEAIKPKKEILEPPEIDLDASMRKLVIVRAGCPIRLFAIVRGPRAPKVTMRKV	19570
QY	102	-----	101
Db	19571	GIDNVRKGQVLDVTMAFLVIPNSTRDSDKYSLTIVNPAGEKAVFVNVRLDTPGPVS	19630
QY	102	-----	101
Db	19631	DLKVSVDTKTSCHVSWAPPENDGGSQVTHYIVKREADRTWSTVTPVKKTSFHVTLNV	19690
QY	102	-----	101
Db	19691	PCNEYFRVTAVNEVPGVPTDVPKVLASDPLSEPPDPKLEVTMTKNSATLAWLPL	19750
QY	102	-----	101
Db	19751	RDGAKIDGYITSYREBEQPADRWTEYSVVKDLSLVVTGLKEGKKYKFRVAARNAGVSL	19810
QY	102	-----	101
Db	19811	PREAGVVEAKEQLLPKILMPEQITIKAGKKLRIEAAHVYKPHPTCKWKKEDEVTTSS	19870
QY	102	-----	101
Db	19871	HLAVHKADSSILIIKDVTBKDSGYSLTAENSGTDTQIKVYVMDAPGPPPPDISD	19930
QY	102	-----	101
Db	19931	IDACSLSWHPILEDGSGNTNIVIVEKDVSGDWLTALASVTKTSRVCKLIPGOEYI	19990
QY	102	-----	101
Db	19991	FRVRAENRFGISEPLTSPKMQAOPFGVPSEPKNARVTKNKDCIFVAMDRPDSGSGPI	20050
QY	102	-----	101
Db	20051	IGYLIERKERNLLWVKANDTLVRSTEYPCAGLVEGLEYSFRIYALNKAAGSSPPSKPTEY	20110
QY	102	-----	101
Db	20111	VTARMPVPPGKPEVIDVTKSTVSLIWARPKHDGSGKIIGYFVEACKLPGBKVRCNTAP	20170
QY	102	-----	101
Db	20171	HQIQOEYATATGLEEKAQVFRAIARTAVNISPPSESDPVTILAENVPPIRDLVAMKS	20230
QY	102	-----	101
Db	20231	LLTVKAGTNVCLDATVFGKPMPTVSKKDGTLCLKPAEGIKWAMQORNLCITLFLSVNRKDS	20290
QY	102	-----	101
Db	20291	GDYTTAENSSGKSATIKLVLDKPGPPASVKINKMYSDRAWLSWEPPLEDGSGSEITNY	20350
QY	102	-----	101
Db	20351	IVDKRETSRPNQAQVSATVPITSCSVEKLIIEGHEQFICAENKYGVSDPVFTEPAIAKN	20410
QY	102	-----	101
Db	20411	PYDPPGRCDDPPVISNITTKOHTMTVSWKPPADGSGPITGYLLEKRETOAQVNTKYNRKPII	20470
QY	102	-----	101

Db	20471	ERTLKATGLQEGTEYEFVRVTAINKAGPGKPSDASKAAYARDPOYPAPPAPPKVYDITTRS	20530
QY	102	-----	101
Db	20531	SVLSWGKPAYDGGSPIIGYLVKVRADSDNWRNCLPNQLOKTRFEVTGLMEDTOYQFR	20590
QY	102	-----	101
Db	20591	VYAVNKIGYSDPSDPDKHPKPDILIPPEGELDADLRKTLILRAGVTMRLVVPVKGRPPP	20650
QY	102	-----	101
Db	20651	KITHSKPNVLRDRIGLDIKSTDFDTFLRCENNVNKYDAGKYILTLENSCGKKEYTIVVKV	20710
QY	102	-----	101
Db	20711	LDTGPPVNVTVKEISKDSAYVTWEPIIDGSPPIINVVKRDAERKSNSTVTTECSKT	20770
QY	102	-----	101
Db	20771	SFRVANLEBGKSYFPRVFAENEYGIGDPGETRDVAKASOTPGPVVDLKVRSVSKSCSIG	20830
QY	102	-----	101
Db	20831	WKKPHSDGSGRIIGYVDFLTEENKQRMKSLSLQYSAKOLTEGKEYTFRVSANENGE	20890
QY	102	-----	101
Db	20891	GTPSBITVVARDDVVAPODLKGLDPLCYLAKENSFRLKPIKIKGPAPSVSWKKGEDPL	20950
QY	102	-----	101
Db	20951	ATDTRSVSESSAVNTLIVYDCQKSDAGKYTITLKNVAGTKEGTISIKVVGKPGIPTGPI	21010
QY	102	-----	101
Db	21011	KFDEVTAEAMTLKWAAPPKDDGGSEITNVILEKRDVSNKNKWTVCASAVOKTTFPRVRLHEG	21070
QY	102	-----	101
Db	21071	MEYTFRVAENKYGVGEGLKSEPIVARHPFDVPDAPPPNIVDVHRHDSVSLTWTDPKKTG	21130
QY	102	-----	101
Db	21131	GSPITGYHLEFKERNLLWKRANKTPIRMDFKVTGLTEGLEYSFVRVMAINLAGVKPSL	21190
QY	102	-----	101
Db	21191	PSEPVALDPIDPPGKPEVINITRNSVTLIWTEPKYDGGHKLGTGYIVEKGLPSKSWMKA	21250
QY	102	-----	101
Db	21251	NHNVPECAFTVTDLVEGGKYEFIRAKNTAGAISAPSESTETIICKDEYEAPTIVLDP	21310
QY	102	-----	101
Db	21311	IKDGLTIKAGDTIVLNAISILGKPLPKSSWSKAGKDIRPSDITQITSTTSSMLTIKYAT	21370
QY	102	-----	101
Db	21371	RKDAGEYTTATNPPFGTKVEHVKVTVLDVPGPPGVEISNVSAEKATLTWTTPLEDGGSP	21430
QY	102	-----	101
Db	21431	IKSVILEKRETSRLMTVWSEDIQSCRHVATKLIQNEVIFRVSANVHYKGEVPOSEP	21490
QY	102	-----	101
Db	21491	KWDRFGPPGPEKPEVSNVTKNATVSWKRPVDDGSGEITGYHVERREKSLRWRAIK	21550
QY	102	-----	101

Db	21551	TPVSLRCKVTGLQEGSTYEFVRVSAENRAGIGPPEASDSVLMKDAAYPPGPPSPHPTD	21610
Qy	102	-----LYITRF-----	107
Db	21611	TTKKSASLAWKPHYDGGLEITGYVVEHQVGDEAWIKDTTGTALRIITQFVVPDLQTKXK	21670
Qy	108	-----	107
Db	21671	YNFRISAINDAGVGEPAVDPVEIVEREMAPDFELDAELRRLTVVRAGLSIRIFVPIKGR	21730
Qy	108	-----	107
Db	21731	PAPEVTWTKDNINLKNRANIENTESFTLLIIPECNRYDTGKEVMTIENPAGKKSGFVNR	21790
Qy	108	-----	107
Db	21791	VLDTPGVLNLRPTDITKDSVTLHMDPLIDGGSRTINIVIEKREATRKSYSTATTKCHK	21850
Qy	108	-----	107
Db	21851	CTYKVTGLSEGCEYFPRVMAENEYIGIGETPETTEPVKASEAPSPDLSNIMDITKSTVSL	21910
Qy	108	-----	107
Db	21911	AWPKPKHGGSKITGYVIEAQRKGSQDWITHITTVKGLECVNRNLTEGEEYTFQVMVNSA	21970
Qy	108	-----	107
Db	21971	GRSAPRESRPVKEQTMPELDRGIYOKLVIKAGDNIKVEIPVLRPKPTVTWTKGD	22030
Qy	108	-----	107
Db	22031	QILKQORVNFETATSTILNINECVRSDSGPYPLTARNIVGEVDVITIQVHDIPGPPT	22090
Qy	108	-----	107
Db	22091	GPIKFEVSSDFVTFSDPDPDGGVPISNYVVMRQDSTTWVELATTVIRTYKATRL	22150
Qy	108	-----	107
Db	22151	TTGLEQYQVRKAQNRVGVGPGITSACIVANYPFKVPGPPTQVAVTKDSMTISWHEPL	22210
Qy	108	-----	107
Db	22211	SDGSPILGYHVERKERNGILWQVSKALVPGNIFKSSGLTDGIAYBFRVIAENMAGSK	22270
Qy	108	-----	107
Db	22271	PSKPSPEMLALDPIDPPCKVPPLNITRHTVTLKWAPEYTGFGKITSYIIVEKRDLPNGRW	22330
Qy	108	-----	107
Db	22331	LKANFSNILENFTVSLGTEDAAYEFVRVIAKNAAGAISSPSEPDAITCRDDVEAPKIV	22390
Qy	108	-----	107
Db	22391	DVKFKDTVLKAGEAFRLEADVSGRPPTMEWSKDGELEGTAKLEIKIADFSTNLVND	22450
Qy	108	-----	107
Db	22451	STRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPEGLAVTEVTSEKCVLSWPPPLDDG	22510
Qy	108	-----	107
Db	22511	GAKIDHYIVQKRETSRLAWNVASEVQVTKLVTKLKNGEYIFRVMVANKYGVGEPLS	22570
Qy	108	-----	107
Db	22571	EPVLAVNPYPPDPKNPEVTTITKDSMVVCMGHDPDGGSEIINYIVERRDKAGQRIWK	22630
Qy	108	-----	107
Db	22631	CNKKTLDLRYKVSLTEGHEGYEFIRMAENAGISAPSTSPFYKACDVTFKPGPPGNR	22690
Qy	108	-----	107
Qy	108	-----	107
Db	22691	VLDTRSSISIAWKNPIYDGSBITGYMVEIALPEEDHQIVTPPAGLKATSYTITGLTE	22750
Qy	108	-----	107
Db	22751	NOBYKIRIYAMNSEGELGEPALVPGTPKAEDRMPLPPEILDADLRKVVITIRACCTLRLFVP	22810
Qy	108	-----	107
Db	22811	IKGRPAPEVKWARDHGESLDKASIESSTSYLLIIVGNVNRFDGKYILTIVENSNGSKSAF	22870
Qy	108	-----	107
Db	22871	VNVRVLDTPGPPDLKVKEVTKTSVTLTWDPPLLDGGSKIKNYIVEKRESTRKAYSTVAT	22930
Qy	108	-----	107
Db	22931	NCHKTSWKVDOLQEGCSYFRVLAENEYIGLPAETAESVKASERPLPPGKITLMDVTRN	22990
Qy	108	-----	107
Db	22991	SVSLSWEKPEHGGSRILGYIVEMQTKGSKWATCATVKVTEATITGLIOGEEYSFRVSA	23050
Qy	108	-----	107
Db	23051	QNEKGISDPRLQSVPIAKDLVIPAPAKLLPNTFTVLAGEDLKVDPFPIGRPTPAVTWHK	23110
Qy	108	-----	107
Db	23111	DNVPLKQTRVNAESTENNLSLLTIKDACREDVGHVVKLTNSAGEAETLNVILDKPGP	23170
Qy	108	-----	107
Db	23171	PTGPKMDEVTAADSTILSWGPKYDGGSSINNYIVEKRDSTTTWQIVSATVARTTIKAC	23230
Qy	108	-----	107
Db	23231	RLKTCGYOFRIAAENRYGKSTYLNSEPTVAQYFPKVPGPPTPVVTLSSRDSMEVOWNE	23290
Qy	108	-----	107
Db	23291	PISDGSRVIGYHLERKERNSTLWVKNKTRIPQTKFTTGLEGEVGEYFRVSAENIVGI	23350
Qy	108	-----	107
Db	23351	GKPSKVSECYVARPCDPPGRPEAIIVTRNSVTLOWKKPTYDGGSKIICYIVEKKELPEG	23410
Qy	108	-----	107
Db	23411	RMWKASFTNIIDTHFEVTLVEDHRYEFRVIAARNAAGVPSESESTGAIARDEVDPRI	23470
Qy	108	-----	107
Db	23471	SMDPKYKDTIVVHAGESFKVDADYCKPIPTIQWIKGDELNTARLEIKSTDFATSLV	23530
Qy	108	-----	107
Db	23531	KDAVRVDSGNYILKAKNVAGERSVTNVKVLDRPGPPEGPVVISGVTAEKCTLAWKPPLO	23590
Qy	108	-----	107
Db	23591	DGSDIINYIVERRETSRLVWTVVDANVQTLCKVTUKLLEGNEYTFRIMAVNKYGVGEPL	23650
Qy	108	-----	107
Db	23651	ESEPVVAKNPFVVPDAPKAVEVTTVTKDSMIVVWERPASDGGSEILGYVLEKRDKEGIRW	23710
Qy	108	-----	107
Db	23711	TRCHKRLIGELRLRVTLGTENHDYEFVRVSAENAGLSEPPSAYQKACDPIYKPGPPNN	23770

Qy 108 ----- 107
Db 23771 PKVIDITRSSVFLWSKPIYDGGCEIQYIVERKCDVSVGEWMTCTPTTGINKTNI EVEKL 23830
Qy 108 ----- 111
Db 23831 LEXHEYFRICAINKAGVGEHADVPGPPIVEEKLEAPDIDLDLELRKIINIRAGGSLRF 23890
Qy 112 ----- 111
Db 23891 VPIKGRPTPEVKMGKVDGEIRDAAIIDVTSSFTSLVDNVNRYDSGKYTLTLENSGOTS 23950
Qy 112 ----- 111
Db 23951 AFVTVRLDTPSPVNLKVTEITKDSVSITWPEPLDGGSKIKNYIVEKREATRKSAAV 24010
Qy 112 ----- 111
Db 24011 VTNCHKNSWKIDLOEGCSYFRVTAENEYIGLPAQTADPIKVAEVPQPPGKITVDVDT 24070
Qy 112 ----- 111
Db 24071 RNSVSLSWTKPEHGGSKIQYIVEMOAKHSEKWCARVKSLOAVITNLTOGEEVLFV 24130
Qy 112 ----- 111
Db 24131 VAVNEKGRSDPRSLAVPIVAKDLVIEPDVKPAFSSYVQVQODLUKIEVPISGRPKPTITW 24190
Qy 112 ----- 111
Db 24191 TKDGLPLKQTRINVTDSLDTLTSIKETHKDDGGQYITVANVVGOKTASIEIVTLDPK 24250
Qy 112 ----- 111
Db 24251 DPPKGVKFDVSAESITLSNPLTYGGCOITNVIQKRDTTTTVMDVSVATVARTLK 24310
Qy 112 ----- 111
Db 24311 VTKLKTGYOFRIFAENRYGQSALES DPIVAQPYKEPGPGPTPATAISKDSMWIQW 24370
Qy 112 ----- 111
Db 24371 HEPVNGGSPVIGYHLERKERNILWTKVNTIIHDTQFKAQNEEGIEYEFVVAENIV 24430
Qy 112 ----- 111
Db 24431 GVGKASKNSECYVARDPCDPPGTPPEIMVXRNEITLOWTKPYDGGSMITGYIVEKRDLP 24490
Qy 112 ----- 111
Db 24491 DGRMMKASFTNVIQFTVSLTEDQRYEFVRVIAKNAAGAIKSPSDSTGPITAKDEVELP 24550
Qy 112 ----- 111
Db 24551 RISMDPKFRDITVNAVGETFRLEADVHGKPLPTIEMLRGDKEIESARCEIKNTDFKALL 24610
Qy 112 ----- 122
Db 24611 IVKDAIRIDGGQYILRASNAVAGSKFPVNVKVLDRPPEGPVQVTGVTSEKSLTWSPP 24670
Qy 123 -OLGG- ----- 126
Db 24671 LDGGSDISHYVVEKRETSRLANTVVASEVVTNSLKVTKLLLEGVEYVFRIMAVNKYGVGE 24730
Qy 127 ----- 126
Db 24731 PLESAPVLMKNPFLPGPPKSLVNTIAKDSMTVCWNRPSDGGSEIIIGYIVEKRDRSGI 24790
Qy 127 ----- 130
Db 24791 RWIKCNKRITDLRLVTLGTEDHEYEFVRSABNAAGVGPSPATVYKACDPVFKGPP 24850
Qy 131 ----- 130

Db 24851 TNAHIVDTTKNSITLAWGKPIYDGGSEILGVVVEICKADEEEOIIVTPQTGLRTRFEIS 24910
Qy 131 ----- 130
Db 24911 KLTEHQEYKIRVCALNKVGLGEATSVPGTVKPEDKLEAPELDOLSELKGIIVRAGGSAR 24970
Qy 131 ----- 133
Db 24971 HIIPKGRPTPEITHSRREGFTDKVQIEKGVNYTQLSIDNCDRNDACKYILKLENSGSS 25030
Qy 134 ----- 133
Db 25031 KSAFVTVKVLDTGPPQNLAVKEVRKDSAFVMEPPIIDGGAKVKVYVIDKRESTRKAYA 25090
Qy 134 ----- 133
Db 25091 NVSSKCSKTSFKVENLTGAIYYFRMAENEGVGPVETVDVAKAAEPSPPGKVTLTD 25150
Qy 134 ----- 133
Db 25151 VSQTSASLMWEKPEHGGSRVLGVVEMQPKOTEKMSIVAESKVCNAVVTGLSSGOEQYOF 25210
Qy 134 ----- 133
Db 25211 RVKAYNEKGS DPRVLGVPVIAKDLTIQPSLKLPTNTYSIOAGEDLKIEIPVIGRRPNI 25270
Qy 134 ----- 133
Db 25271 SMVKDGEPLKQTRVNVBEATSTVLHIKEGNKDDFGKTVTATNSAGTATENLSIVILE 25330
Qy 134 ----- 133
Db 25331 KCPVPVGRFDEVSADFFVISWEPAYTGGCOISNYIVEKRDTTTTTHMVSVATVART 25390
Qy 134 ----- 133
Db 25391 IKITLKTGYOFRIFAENRYKGSAPLDSKAVIQVPPKEPGPGTPTPTSISKDOMLV 25450
Qy 134 ----- 133
Db 25451 QMHEPVNDGGTKIIGYHLEQEKNSILWVKLNKTPIQDTKFTKTGLDEGLEVEFKVSAEN 25510
Qy 134 ----- 133
Db 25511 IVGIGKPSKVSECFVARDPCDPPRDEAIVITRNNVTLKWKPAYDGGSKITGYIVEKKD 25570
Qy 134 ----- 133
Db 25571 LPDGRWKASFTNVLTEFTVSLVEDQRYEFVRVIAARNAAGNPFSEPSDSQAITARDEID 25630
Qy 134 ----- 133
Db 25631 APNASLDPKYKDVIVVHAGETFVLEADIRGKPIPDVVMVSKDGELEBETAARMEIKSTIQK 25690
Qy 134 ----- 133
Db 25691 TTLVWKDCIRTDGGQYILKLSNVGKTSIPITVKVLDLPPEGLKVTGVTAEKCYLAW 25750
Qy 134 ----- 133
Db 25751 NPPLQDGGANISHYIEKRETSRLSWTQVSTEVOALNYKVKLLPGNEYIFRVMVAVNKY 25810
Qy 134 ----- 133
Db 25811 IGEPLESGPVTACNPKYKPPGPPSTPEVSAITKDSMVVTWARPVDDGGTEIEGYILEKRD 25870
Qy 134 ----- 133
Db 25871 EGVWTKCNKKTLDLRLVTLGTGHSYEFVRAAENAGVGPSEPSVYRACDALYPP 25930
Qy 134 ----- 133

Db 25931 GPPSNPKVTDTSRSSVSLAWSKPIYDGGAPVKGYVVEVKEAAADEWTTCTPTTGLQKQF 25990
Qy 134 ----- 133
Db 25991 TVTKLKENTYFNFRICAINSEGVEGEPATLPGSVVAQERIEPPEIELDADLRKVVVL RASA 26050
Qy 134 ----- 133
Db 26051 TLRLFTVTKRPEPEVKEAEGILTDRAQIEVTSFTMLVIDNVTRFDGRYNLTLENN 26110
Qy 134 ----- 133
Db 26111 SGSKTAFVNVRLDSPAPVNLTIREVKKDSVTL SWEPPLEDGGAKITNYIIVEKETTRK 26170
Qy 134 ----- 133
Db 26171 AYATINNCTKTFRIENLEQEGSYFRVLASNEYGIGLPAETTEPVKVEBPPLPPGRVT 26230
Qy 134 ----- 133
Db 26231 LVDVTRNTATIKWKPESDGGSKITGYVVMOTKGSEKWSCTQVKTLTATISGLTAGEE 26290
Qy 134 ----- 133
Db 26291 YVFRVAANEKGRSDRQLGCPVIARDIEIKPSVELPFHTFNVKAREQLKIDVPFKGRPQ 26350
Qy 134 ----- 133
Db 26351 ATVNRKDGOTLXETTRVNVSSKTVTSLSIKEASKEDVGTLYELCVSNSAGSITVPITII 26410
Qy 134 ----- 133
Db 26411 VLDRPGPPGPIRIDEVSCSITISWNPPEYDGGCQISNYIIVEKETTSTWHIVSOAVAR 26470
Qy 134 ----- 133
Db 26471 TSIKIVRLTTGSEYQFRVCAENRYGKSSYSESSAVAEYFPSPGPGTPKVVHATKSTM 26530
Qy 134 ----- 133
Db 26531 LVTWQVPNDGSRVIGVHLEYKERSILWSKANKILIAOTOMKVSGLDEGLMYEYRVA 26590
Qy 134 ----- 133
Db 26591 ENIAGICKSKSEVPAPDPCDPGQPEVNTIRKSVLSLWSPKPHYDGGAKITGIVER 26650
Qy 134 ----- 133
Db 26651 RELPDGRWLKCNVTNIQETYPEVTELTEDORYEFRVFARNAADSVSPSESTGP11VKDD 26710
Qy 134 ----- 133
Db 26711 VEPFRVMDVKFRDVIIVKAGEVLKINADIAGRPLPVISWAKOGIEIERARTEIISTDN 26770
Qy 134 ----- 133
Db 26771 HTLLTVKDCIRDTQGYVLT1KNVAGTRSAVNCVKLDKPPAGPLEINGLTAEKCSLS 26830
Qy 134 ----- 133
Db 26831 WCRPOEDGADIDYIIVEKRETSHLAWTICEGELQMTCKVTKLLKGNEYIFRVTGVNKY 26890
Qy 134 ----- 133
Db 26891 GVGEPLESVAIKALDPFTVPSPPTSLEITSVTKESMTLCWSRPESDGGSEISGVIIERRE 26950
Qy 134 ----- 133
Db 26951 KNSLRWVRNKKPVYDLRVKSTGLREGCEYEVYVAENAGLSLPSETSPLIRAEDPVFL 27010
Qy 134 ----- 133
Db 27011 PSPPSKPIVDGKTTITIAWVKPLFDGGAPITGYTVVEYKKSDDTDWKTISIQSLRGTYT 27070

Qy 134 ----- 133
Db 27071 ISGLTTGAEEVFRVKNVKGASDPSSDPOIAKEREBEPLFDIDSEMRKTLIVKAGAS 27130
Qy 134 ----- 133
Db 27131 FTMVTPFRGRPVNVLWSKPDOTLRTAYVDDTTDSRTSLTIENANRNDOSKYTLTIONVL 27190
Qy 134 ----- 133
Db 27191 SAASLT1LVVKVLDTPGPPPTNITVODVTKESAVLSWDVPENDGGAPVKYNYHIEKREASKA 27250
Qy 134 ----- 133
Db 27251 WSVTNNCNRLSYKVTLQEGAIYVFRVSGENEFVGIPAETKEGVKITEKPSPEPLGV 27310
Qy 134 ----- 133
Db 27311 TSISKDSVSLTWLKPHEHGGSR1VHYVVEALEKGQKNWCAVAKSTHHVVVSGLRENEY 27370
Qy 134 ----- 133
Db 27371 FFRVFAENQAGLSDPRELLLPVLIKEQLEPPEIDMKNFPSSHVVYVVRAGSNLKVDPISGK 27430
Qy 134 ----- 133
Db 27431 PLPKVTLSDRGVPLKATMRFNTEITAENLTLINKESVTADAGRYEITAAANSSGTTKAFIN 27490
Qy 134 ----- 133
Db 27491 IIVLDRPGPTGPVVISDITEESVTLKWEPPKYDGGSOVTHYILLKRETSTAVMTVEUSAT 27550
Qy 134 ----- 133
Db 27551 VARTMKVMKMLTTGEEYQFRIKAENRFGISDHIDSACVTVKLPTVTPGPPSTPMVTVNTR 27610
Qy 134 ----- 133
Db 27611 ESITVWHEPVSNGGSAVVGHYHLEMKDRNSILWOKANKLVIRTHFKVTTISAGLIYEFR 27670
Qy 134 ----- 133
Db 27671 VYAENAGVGKPSHPSEPVLAIDACEPPRNVIRITDISKNSVLSLWQQPAFDGSGKITGYI 27730
Qy 134 ----- 133
Db 27731 VERRDLPGRWTKASFNTVETQFIISGLTQNSQYEFVRVARNVAGSISNPSEVVGPIFC 27790
Qy 134 ----- 133
Db 27791 IDSYGGPVIDLPLEYEVVVKYRAGTSVKLRAGISGKPAPTIEWYKDDKELOTNALVCVEN 27850
Qy 134 ----- 133
Db 27851 TTDLASILIKDADRLNSGCYELKLRNMGASATIRVOILDKPPGPGPIEFKTVTABKI 27910
Qy 134 ----- 137
Db 27911 TLLWRPPADGGAKITHYIIVEKRETSRVVWVSMVSEHLEECIIITTTKIIKGNEYIFRVRV 27970
Qy 138 ----- 137
Db 27971 NKYGIGEPLESDSVAKNAFVTPGPGPIEVTKITKNSMTVWWSRPIADGGSDISGYFLE 28030
Qy 138 ----- 137
Db 28031 KRDKSLGWFKVLKETIRDTROKVTGLTENS DYQYRVCAVNAAGOGPFSEPFYKAADP 28090
Qy 138 ----- 137
Db 28091 IDPPGPKIRIADSTKSSITLGMSPVYDGGSAVTGYVVEIROGEBEETVSTTSGEVR 28150

Qy	138	-----	137
Db	28151	TTEVVSNLXPGVNYIFRVSANVCAGGEPENNEPVQAKDILEAPEIDLVALRTSVIA	28210
Qy	138	-----	137
Db	28211	KAGEDVOQLIPFKGRPPPTVTRKDEKNLGDARYSIENTDSSLLTIPQVTRNDTGKYI	28270
Qy	138	-----	137
Db	28271	LTINGVGEPSKSTVSXKVLDTPAACOKLQKHVSRGTVTLWDPLIDGSPFIINXVIE	28330
Qy	138	-----	137
Db	28331	KRDATKRTWSVUSHKCSSTSFKJLDLSEKTPFFPRVLAENEIGEGPCEETEPVKAAEVP	28390
Qy	138	-----	137
Db	28391	APIRDLNMDKSTKTSVILSWTKPDPDGGSVITEYVVERKKGEGTWSHAGISKTCEIEVS	28450
Qy	138	-----	137
Db	28451	OLKEQSVLEFRVPAKNEKGLSDPVTIGPITVKELIITPEVDLSOIPCAQVTVRIGHNVHL	28510
Qy	138	-----	137
Db	28511	ELPYKGRPKPSISMLKDGLPLKESEFVRFSKTENKITLSIKNAKKEHGKYTVILONAVC	28570
Qy	138	-----	137
Db	28571	RIAVPITVITLGPSPKPKPIRFDKADSVILSWDPEDNGGGEITCYSIKRETSQTN	28630
Qy	138	-----	137
Db	28631	WMVCSSVARTTFKVPNLVKDAEYQFVRAENRYGVSOPLVSSIIVAKHOFIRPGPGKP	28690
Qy	138	-----	137
Db	28691	VIYNTSDCMLTWDAPVYDGGSEVTGPHVEKKERNISILMOKVNTSPISGREYRATGLVE	28750
Qy	138	-----	137
Db	28751	GLDYQFRVYAENSAGLSSPDSKFTLAVSPVDPPTDYPDYIDVTRETITLKNPPLRDGG	28810
Qy	138	-----	137
Db	28811	SKIVGSIEKROGNERWRCNFTDVSECYTVTGLSPGDRVEFRIIARNAVGTISPSPQS	28870
Qy	138	-----	137
Db	28871	SGIIMTRDENVPPIVEFGPEYFDGLIIKSGESLRIKALVQGRPVPRVTFWKDGVIEIKRM	28930
Qy	138	-----	137
Db	28931	NMEITDVLGSLFVRDATRDHRGVYTVKAKNAGSAAEIKVKQDTPGKVVGPIRFTN	28990
Qy	138	-----	137
Db	28991	ITGEKWTLMWADPLNDGCAPITHYIIEKRETSRLAWALIEDKCEAQSYTAIKLINGNEYQ	29050
Qy	138	-----	137
Db	29051	FRVSANVKGVRPLSDPWAQIQYTVDPAPGIPSPNITGNSITLTWARPESDGGSEI	29110
Qy	138	-----	137
Db	29111	QQVILERREKSTRVVKVISKRPISETRFKVTGLTEGNEYEFHVMAENAGVGPASGISR	29170
Qy	138	-----	137
Db	29171	LIKREPVNPPGPTVVKVTDTSKTTVSLEWSKPVFDGGMGIIGYIEMCKADLGDWHKV	29230
Qy	138	-----	137

Db	29231	NAEACVKTRYVTDLQAGEYKFRVSAINGAGKDSCEVTGTIKAVDRLTAPELDIDANF	29290
Qy	138	-----	137
Db	29291	KQTHVVRAGASIRLFIAYQGRPTPTAVMSKPDNSLRAIDIHTTDSFTLTVCNCRNDA	29350
Qy	138	-----	137
Db	29351	GKYLTVNNSGSKSITFTVKVLDTPGPPGPIPTPKDVTGRSATLMDAPLDDGGARIHYY	29410
Qy	138	-----	137
Db	29411	VVEKREASRRSQWVISEKCTRQIFKYNDLAEGVPYFRVSAVNEVGCEPYEMPEPIVAT	29470
Qy	138	-----	137
Db	29471	EQPAPPRLLDVVDTSKSSAVLAWLKPDDHGGSRITGYLLEMRKQSGDFWVBAGHTKQJTF	29530
Qy	138	-----	137
Db	29531	TVERLVEKTEYEFVRVAKNDAGYSEPREAFSVIIKEPQIEPTADLTGITNQLITCKAGS	29590
Qy	138	-----	137
Db	29591	PFTIDVPISGRPAPKVTKLEMLKETDRVSIITTKDKORTTLTVKDSMRGDSGRYFLTLE	29650
Qy	138	-----	137
Db	29651	NTAGVKTFSTVVJIRPGPVTGPTEVSSVSAESCULSWGEPKDDGGTEITNIVYVEKRES	29710
Qy	138	-----	137
Db	29711	GTTAMOLVNSVVKRTQIKVTHLTKYMEYSFRVSSSENRFGVSKPLESAPIAIEHPVPESA	29770
Qy	138	-----	137
Db	29771	PTRPEVYHVYANAMSIRWEEPYHDGSKIYGVWEKERNTILWVKENKVPCECNYKYVT	29830
Qy	138	-----	137
Db	29831	GLVEGLEQFRTYALNAGVSKASEASRPIMAQNPVDAPGRPEVTDVTRSTVSLIWSAPA	29890
Qy	138	-----	137
Db	29891	YDGGSKVGYIIERKPVSEVGDGRWLKCNVTIVSONFTVTALSEGDTYEFVRLAKNAAG	29950
Qy	138	-----	137
Db	29951	VISKGSESTGPVTCRDEYAPPAELDARLHGLVTIRAGSLVLDAAVQKPEPKIIMTK	30010
Qy	138	-----	137
Db	30011	GDKELDLCEKVSLOYTGKATAVIKFCDRSDSGKYTLTVKNASGTVKAVSNVUKVLDSPG	30070
Qy	138	-----	137
Db	30071	CGKLTVSRVTOEKCTLAWSLPQEDGGAETHYIVERRETSRLNWNVIVEGECPTLSYVWTR	30130
Qy	138	-----	137
Db	30131	LIIKNEYIIPRAVAVNYKPGVPVESEPIVARNSTFIPSPGIPPEVGTGKEHIIIOWTKP	30190
Qy	138	-----	137
Db	30191	ESDGGNEISNYLVDKREKKSRLRTRVKNKYVVYDTRLKVTLMEGCDYQFRVTVAVNAGN	30250
Qy	138	-----	137
Db	30251	SESEASNFISSCREPSYTPGPSPAPRVVDTTKHSISLAWTKPMYDGGTDIVGVYLEMOEK	30310
Qy	138	-----	137

```
Db 30311 DTDQYRVHTNATIRNTEFTVPDLKXGQKYSFRVAANVXGMSXSESIABIEPVERIEI 30370
Qy 138 ----- 137
Db 30371 PDLELADDLKKTVTIRAGASLRMLVSGRPPPVITWSKQIDLASRAIIDTTESYSILLI 30430
Qy 138 ----- 137
Db 30431 VDKVNRDAGKYTIEAENQSGKSATVLVKVYDTFGPCSPSVKVEVSRDSVTITWIEIPTI 30490
Qy 138 ----- 137
Db 30491 DGGAPNNYIVEKREAAAFKTVTTKCSKTLRISGLVEGMYFVRLPENIYIGGPC 30550
Qy 138 ----- 137
Db 30551 ETSDAVLVSEPLVPAKLEWVDVTKSTVTLAMEKPLYDGGSLRTGYVLEACKAGTERWNK 30610
Qy 138 ----- 137
Db 30611 VVTLKPTVLEHTVTSNEGEQVLFRIARQNEKGVSEPRETVTAVTVQDLRVLPITDLSM 30670
Qy 138 ----- 137
Db 30671 POKTHVPACRPVELVPIAGRPPPAASWFFAGSKLRESERVTVETHTKVAKLTIRETTI 30730
Qy 138 EDTG----- 141
      III
Db 30731 ROTGEVTELEKNVTGTTSETIKVILDKPGPTGPIKIDEIDATISITISWEPPELDGAP 30790
Qy 142 ----- 141
Db 30791 LSGYVVEQRDAHRPGWLPSVESVTRSTFTKPTRLTEGNEYVFRVAATNRFIGSYLQSEVI 30850
Qy 142 ----- 141
Db 30851 ECRSSIRIPGPPTLOIFDVSRDGMTLWYPPEDDGSQVTGYIVERKEVRADRWVRNK 30910
Qy 142 ----- 141
Db 30911 VPTMTRYRSTGLTEGLEVEHRTAINARGSKPSRSPKPIVAMDPIAPGPKPQNPRVTD 30970
Qy 142 ----- 141
Db 30971 TTRTSVSLAWSPEDEGSKVTGYLIEMQKVDQHEWTKNTTPTKIREYTLTHLPOGAEV 31030
Qy 142 ----- 141
Db 31031 RFRVLACNAGGPEAEVPGTVKVTMLEYDPDYELDERYQEGIFVRQGGVIRLTIPK 31090
Qy 142 ----- 141
Db 31091 PFPICKWTEGQDISKRAMIATSETHTELVIKEADRGDSGTVDLVLENKCGKKAIVIKVR 31150
Qy 142 ----- 141
Db 31151 VIGSPNPEGLEVDYDIQVRSVRVSWRPPADGGADILGYILERREVPKAAWVTIDSRVR 31210
Qy 142 ----- 141
Db 31211 GTSLVVKGLENVEYHFRVSAENQFGISKPLKSEEPVTPKTLNPPPEPSNPPEVLDVTK 31270
Qy 142 ----- 141
Db 31271 SSVSLWSRPPKDDGSGRVGTGYIERKETSTDKVVRHNKTQITTTMYTVTGLVPAEYQFR 31330
Qy 142 ----- 141
Db 31331 IIAQNDVCLSETSPASEPVVCKDFDXPSQPELEILSISKDSVTLQWKEPCDGGKEIL 31390
Qy 142 ----- 141
Db 31391 GYWVEYRQSGDSAWKSKNERIKDKQFTIGLLEATEYEFVRFAENETGLSRPRRTAWSI 31450
```

```
Qy 142 ----- 141
Db 31451 KTKLTSEAPGIRKEMKDVTTKLGEAAQLSCQIVGRPLDKWYRFKELIOSRKYKMS 31510
Qy 142 ----- 141
Db 31511 DGRTHTLTVMTEQDEBEGVYTCIATNEVGEVETSSKLLQATQFHPGYPKKEKYCAGV 31570
Qy 142 ----- 141
Db 31571 STLRLHVMYIGRPVPAMTWFGOKLLQNSENIENTEHTYHLVMKNVORKTHAGKYKV 31630
Qy 142 ----- 141
Db 31631 LSNVFGTVDAILDVEIQDKPKTGPVIVIBALLKNSAVISWKKPPADGGSMITNYVVEKC 31690
Qy 142 ----- 141
Db 31691 EAKEGAEWQLVSSAISVTTCTRIVNLTENAGYFVRSAQNTFGISDPLEVSVVLIKSPFE 31750
Qy 142 ----- 141
Db 31751 KPGAPGPTTAVTKDSCVVAWKKPPADGGAKIRNYYLEKREKKONKMWISVTTEEIRETV 31810
Qy 142 ----- 141
Db 31811 FSVKNLTIEGLEVEFRVKENLGGESEWSEISEPITPKSDVPIOAPHKPEELRNLNRYQS 31870
Qy 142 ----- 141
Db 31871 NATLVCKVTGHPKPIVKYRQGEIADGLKYRIOEFGYHQHLLIASVTDVDDVTVQVR 31930
Qy 142 ----- 141
Db 31931 ATNOGGSVGTASLEVEVPAKIHLPKTLEGWGAVALRGEVVSIKIPSGKPDVITWQK 31990
Qy 142 ----- 141
Db 31991 QODLIDNNGHYQVIVTRSFSLVPNGVERKDAGFYVVCANRFGIDOKTVELDVADVPD 32050
Qy 142 ----- 141
Db 32051 PPRGVKVSVDVSRDSVNLTWTEPASDGGSKIINYIVEKCATTAERWLRVGOARETRYVIN 32110
Qy 142 ----- 141
Db 32111 LFGKTSYQFRVIAENKFGLSKSESEPTITKEDKTRAMNYDEEVDRETVSMTKASHSS 32170
Qy 142 ----- 141
Db 32171 TKELYEKYMIABDLGRBFGIVHRCVETSCKKTYMAKFKVKGTDQVLVKKEISILNIAR 32230
Qy 142 ----- 141
Db 32231 HRNIIHLHESFESMEELVMIPEFISGLDIFERINTSAFELNREIVSVYHVQCEALQFLH 32290
Qy 142 FDL----- 144
      III
Db 32291 SHNIGHFDIRPENIIVQTRRSSTIKIIBFGOARQLKPGDNFRLFTAPEYAPAEVHQHVD 32350
Qy 145 ----- 144
Db 32351 VSTATDMMSLGLTVYVLLSGINPFLAETNQIINIMINAEYTFDEAPKEISIEAMDFFVD 32410
Qy 145 ----- 144
Db 32411 RLLVKERKSRMTASEALQHPWLKQKIERVSTKVIRTLKHRRYYHTLIKKDLNMVWSAARI 32470
Qy 145 ----- 144
Db 32471 SCGATRSQKGVSAKVAKVASIEIGVSGQIMHAVGEGGHVKYVCKIENYDQSTQVTWY 32530
```

```
Qy 145 ----- 144
Db 32531 FGVRQENSEKEITYEDGVALLYVKDITKLDDCTYCKVNDYGDSSVAELFVKGVRE 32590
Qy 145 -----CVTI 148
Db 32591 VDYDCRRMKKIKRTRDTMRLRPPPEFTPLYNKYAYGVNVRFGVTI 32640

RESULT 3
Q10466 PRELIMINARY; PRT; 26926 AA.
AC Q10466
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Titin, heart isoform N2-B (EC 2.7.1.-) (Connectin).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RX MEDLINE=96026330; PubMed=7569978;
RA Labelit S., Kolmer B.;
RT "Titins: giant proteins in charge of muscle ultrastructure and
RT elasticity.";
RL Science 270:293-296(1995).
RN [2]
RP SEQUENCE OF 22277-25376 FROM N.A.
RX MEDLINE=92259380; PubMed=1582406;
RA Labelit S., Gautel M., Lakey A., Trinick J.;
RT "Towards a molecular understanding of titin.";
RL EMBO J. 11:1711-1716(1992).
RN [3]
RP SEQUENCE OF 1976-2014 FROM N.A.
RA Labelit S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95331114; PubMed=7607248;
RA Gautel M., Castiglione-Morrelli M.A., Pfuhl M., Motta A., Pastore A.;
RT "A calmodulin-binding sequence in the C-terminus of human cardiac
RT titin kinase.";
RL Eur. J. Biochem. 230:752-759(1995).
CC -1- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE
CC ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF
CC SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE
CC PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE
CC DISTRIBUTION. DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY
CC ONE TISSUE. THE SEQUENCE SHOWN HERE IS THAT OF THE HEART ISOFORM
CC N2-B.
CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
CC -1- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE
CC KINASES.
CC -1- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 112
CC IMMUNOGLOBULIN C2-LIKE DOMAINS AND 132 FIBRONECTIN TYPE III-LIKE
CC DOMAINS.
DR EMBL; X64698; CAA45939.1; -
DR EMBL; X83270; CAA58243.1; -
DR EMBL; X64697; CAA45938.1; -
DR EMBL; X90568; CAA62188.1; -
DR EMBL; X64699; CAA45940.1; -
DR HSP; P56276; 1TLK.
DR InterPro; IPR000282; Cytok receptor_2.
DR InterPro; IPR00719; Euk pKinase.
DR InterPro; IPR005577; FGGY_kin.
DR InterPro; IPR003962; FhIII_repeat.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001092; HLH_basic.
DR InterPro; IPR003598; Ig_C2.
```

```
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000129; Peptidase_S24.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00041; fn3; 132.
DR Pfam; PF00047; Ig; 91.
DR Pfam; PF00069; pkinase.1.
DR PRINTS; PR00014; FNTYPEI11.
DR PRINTS; PR00726; LEXASERPTASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00060; FN3; 127.
DR SMART; SM00408; IGC2; 23.
DR SMART; SM00410; IG_like; 79.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00933; FGGY_KINASES_1; UNKNOWN_1.
DR PROSITE; PS0038; HELIX LOOP HELIX; UNKNOWN_1.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE; PS00435; PEROXIDASE_1; UNKNOWN_1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Muscle protein; Cytoskeleton; Structural protein; Calmodulin-binding;
KW Serine/threonine-protein kinase; Alternative splicing; Repeat;
KW Immunoglobulin domain; Phosphorylation.
FT DOMAIN 1370 1389 4 X 5 AA TANDEM REPEATS OF R-M-S-P-A.
FT DOMAIN 4429 4614 GLU/LYS/PRO/VAL-RICH.
FT DOMAIN 24731 25070 CATALYTIC.
FT DOMAIN 25030 25056 CALMODULIN-BINDING.
FT MOD_RES 1372 1372 PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT MOD_RES 1377 1377 PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT MOD_RES 1382 1382 PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT MOD_RES 1387 1387 PHOSPHORYLATION (BY PDPK) (POTENTIAL).
FT MOD_RES 26171 26171 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 26178 26178 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 26184 26184 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 26190 26190 PHOSPHORYLATION (POTENTIAL).
FT CONFLICT 22277 22277 T -> P (IN REF. 2).
FT CONFLICT 22449 22449 E -> G (IN REF. 2).
FT CONFLICT 22454 22454 T -> Q (IN REF. 2).
FT CONFLICT 23324 23324 S -> L (IN REF. 2).
SQ SEQUENCE 26926 AA; 2993428 MW; DSEEC3254DF5523 CRC64;

Query Match 35.2%; Score 272; DB 4; Length 26926;
Best Local Similarity 0.4%; Pred. No. 10;
Matches 101; Conservative 16; Mismatches 23; Indels 24887; Gaps 28;

Qy 9 ELL----- 11
Db 190 ELLVQGEVEVPKTKTIVSTAQISSEOTRIEKKIEHFDARSIAIVENWIDGAAGQOL 249
Qy 12 ----- 11
Db 250 PKHTPPRIPPKPKSRPTPPSIAAKAQLAQOSPIRHSPPVHRVRAPTSPVRSVSP 309
Qy 12 -----VAVGDPV----- 18
Db 310 AARISTSPISRVSPRLMRKTQASTVATGPEVPPPMKQEGYVASSSEAMRETTLTSTQ 369
Qy 19 ----- 18
Db 370 IRTERWEGRYGVOEQVITSGAAGAAASVSASAAEAATGATGKADKSAAVATV 429
Qy 19 ----- 18
Db 430 VAAVDMARVREPVISAVEQTAQRTTTTAVHIQPAEQVRKEAEKTAIVTKVVVAADKAKEQ 489
Qy 19 ----- 18
Db 490 ELKSRTKEIITTKQEQMHVTHEQIRKETEXTFVPKVVVISAAKAEQETRISSEITKKQKQ 549
Qy 19 ----- 18
```

Db 550 VTQEAIMKETRTKTVVPKVIVATPKVKEQDLVSRGREGITTKREQVQIITOEKMRKEAKTA 609
Qy 19 ----- 18
Db 610 LSTIAVATAKAEOTILRTRETMATRQEQIOVTHGKVDVGKKA EAVATVAAVDQARVR 669
Qy 19 ----- 18
Db 670 EPREPGHLEESAQAOTTLEYGYKERISAAKVAEPPORPASEPHVVPKAVKPRVIOAPSET 729
Qy 19 ----- 18
Db 730 HIKTDDOKGMHISSOIKKTTDLTTERLVHVDKRPRTASPHFTVSKISVPKTEHGYEASIA 789
Qy 19 ----- 18
Db 790 GSAIATLQKELSATSQAOKITKSVKAPTVPKSETRVRAEPTPLPQPPFADTPTDYKSEAG 849
Qy 19 ----- 18
Db 850 VEVKKEVGVSIITGTVREERFEVLHGREAKVTETARPAVEIPVTPPTLVSLGNVTVI 909
Qy 19 ----- FQ----- AHQEDTERYV- 30
||| : : :
Db 910 EGESVTLECHISGYPSPVTWYREDYQIESSIDFOITFQSGIARLMIREAFEDSGRFTC 969
Qy 31 ----- 30
Db 970 SAVNEAGTVSTCYLAVOVSEEFKETTA VTEKFTTTEKRFVESRDVWMTDTSLTEQAG 1029
Qy 31 ----- 30
Db 1030 PGEPAAPFYITKPVVQKLVEGGSVFCQVGNPKPHVYWKSGVPLTTGYRYKVSYNQK 1089
Qy 31 ----- 30
Db 1090 TGECKLVISMTFADDAGEYTI VVRNKHGETSASASLLEADYELLMKSOQEMLYQTQVTA 1149
Qy 31 ----- 30
Db 1150 FVQEPVEGETAPGFVYSEYEKEQALIRKKNAKDTVVVTVVVEDQEPHISSEFERLI 1209
Qy 31 ----- 30
Db 1210 KEIERYIKTTLEELLEEDGEEKNAVDISESEAVESGFDLRKYNRILEGMGVTFHCKMS 1269
Qy 31 ----- 30
Db 1270 GYPLPKIAWKDKRIKHGERYQMDFLQDGRASLRIPVLPDEDEGIYTAFAFNKGNALC 1329
Qy 31 ----- 30
Db 1330 SGKLYVEPAAPGAPTYIPTLEPVSRISLSPRSVSRSPRMSPARMSPARMSPA 1389
Qy 31 ----- 30
Db 1390 RMSPPRLEETDESOLERYKPVFLKPVSPFKLEGANCREFDLKVGRPMPETEFWHDGQ 1449
Qy 31 ----- 30
Db 1450 QIVNDYTHKVIKEDGTOSLIIVPATPSDGEWTVVAQNAGRSSISVILTVEAVEHOVK 1509
Qy 31 ----- LTNANI----- 36
||| : : :
Db 1510 PMFVEKLNKVNKEGSRLEMKVRATGNPNPDIVLKNSDIIVPHKPKIRIEGTKGEAL 1569
Qy 37 ----- 36
Db 1570 KIDSTVSQDSAWYTATAINKAGRDTRCKVNVVEFABEPEPERKLIIPRGTYRAKEIAAP 1629
Qy 37 ----- 36
Db 1630 ELEPLHLRYQBQWEEGDLYDKEKQKPPFFKKLTSRLKRFPAHFECRLTPIPSDPTMV 1689

Qy 37 ----- 36
Db 1690 VEWLHDKPLEAANRLRMINEFVGYCSDYGVAYSRDSGIITCRATNKVGTDHTSATLIVK 1749
Qy 37 ----- GA----- 38
||| : : :
Db 1750 DEKSLVEESOLPEGRKGLORIEELERMAHEGALTGVTTDOKEKQKPDIVLYPEPVRVLEG 1809
Qy 39 ----- 38
Db 1810 ETARPRCRVTGYPOPKVNWYNGQLIRKSKRFRVYDGIHYHLDIVDCKSYDTGEVKVTAE 1869
Qy 39 ----- 38
Db 1870 NPEGVIEHKVLEIQOQREDPRSVLRRAPRPEPFVHVEPGKQLEQVQKVDPRVDTTETKE 1929
Qy 39 ----- ELLR--- 42
||| : : :
Db 1930 VVKLRAERITHEKVPBESEELRSKFKRTEEGYEAITAVELKSRKKDESYELLRKTK 1989
Qy 43 ----- 42
Db 1990 DELLHWTKELTEBEKKALAEKGKITPTFKPKDIELSPSMEAPKIFERIQSOTVQGSDA 2049
Qy 43 ----- Dps----- 45
||| : : :
Db 2050 HFRVRVVGKPDPECEWYKNGVKIERSDRIYWYMPEDNVCELVRDVTAEASIMVKAIN 2109
Qy 46 ----- 45
Db 2110 IAGETSSHAFLVQAKOLITFTQELQDVVAKEKOTMATFECETSEPPFVKVWKYKDGMEVH 2169
Qy 46 ----- 45
Db 2170 EGDKYRMSDRKVFHLSILTIDTSDAEDYSCVLVEDENVKTTAKLI VEGAVVEFVKLEQD 2229
Qy 46 ----- 45
Db 2230 IEVPESYGELECI VSPENIEGKWYHNDVELKSNKYTITSRRGRQNLTVKDVTKEDQGE 2289
Qy 46 ----- 45
Db 2290 YSFVIDGKTKTKLKKKPRPIAILQGLSDQKCEGDI VQLEVKVLSLESVEGVMKDGQEV 2349
Qy 46 ----- 45
Db 2350 QPSDRVHIVIDKQSHMLLIEDMTKEDAGNYSFTIPALGLSTGRVSVSDVITPLKDVN 2409
Qy 46 ----- 45
Db 2410 VIEGTKAVLECKVSPDVTSVKWLNDQIKPDDRQVAIVKGTQKORLVINRTHASDEGYP 2469
Qy 46 ----- 45
Db 2470 KLIVGRVETNCNLSVEKIKIIRGLRDLTCTETQNVVFEVLSHSGIDVLWNFKDEIKPS 2529
Qy 46 ----- 45
Db 2530 SKYKIEAHGKIYKLTVLNNMKDDEGKYTFYAGENMTSGKLT VAGGAISKPLTDOTVAESQ 2589
Qy 46 ----- 45
Db 2590 EAVFECEVANPDSKGEWLBDGKHLPLTNNIRSESGHKRRLIIAATKLLDDIGEYTYKVAT 2649
Qy 46 ----- LGAQFRVHL--VKWV----- 58
||| : : :
Db 2710 SVKGTIYSLRIKNCIAIVDESIVYFRLGRLGASARLHVETVKIIPKPKDVTALENATVAPE 2769

QY 59 ----- 58
 Db 2770 VSVSHDTPVKNFHKSVIEIKPSDKHRLVSEKVKHKLMLQNI SPSDAGEYTA VVGQLECKA 2829
 QY 59 ----- 58
 Db 2830 KLFVETLHITKTMKNI EVPETKTASFECEVSHFNVPFMMKNGVEIEMSEKFI VVGQKL 2889
 QY 59 ----- 58
 Db 2890 HOLIIMNTSTEDSAEYTFVCGNDQVSATLTVTPIMITSMKLDINAEKDTITFEVTVNYE 2949
 QY 59 ----- 58
 Db 2950 GISYKWLKNGVEIKSTDKQWRCKKLTHSLNIRNVHFGDAADYTFVAGKATSTATLYVEA 3009
 QY 59 ----- 58
 Db 3010 RHIEFRKHIKIDKVKLEKRAMFECEVSEPDITVQMKDDOELOITDRIKIOKEYVHRL 3069
 QY 59 ----- 58
 Db 3070 IPSTRMSDAGKYTVVAGGNVSTAKLFVEGRDVRIRSIKKEVOVIEKORAVVEFEVNEDDV 3129
 QY 59 ----- 58
 Db 3130 DAHWYKOGIEINFQVQERHKYVVERRIHRMFISETROS DAGEYTFVAGRNRSSVTLVYNA 3189
 QY 59 ----- 58
 Db 3190 PEPPOVLOEQPVTVQSGKPARFCAMISGRPOP KISWYKEQLLSTGFKFLHDGQEYT 3249
 QY 59 ----- 58
 Db 3250 LLLIEAFPDAAYTCEAKNDYGVATTSASLSVEPVSPDQEMPVYPPIITPLQDTV 3309
 QY 59 ----- 58
 Db 3310 TSEGQPARFOCRVSGTDLKVSWSXDKKI KPSRFRMTQFEDTVQLEIAEAYPEDEGTYT 3369
 QY 59 ----- 58
 Db 3370 FVANNAGOVSTANLSLEAPESILHERIEQEIEMEMKEFSSFLSABEEGLHSAELQLS 3429
 QY 59 ----- 58
 Db 3430 KINETLLESVPVYPTKFDSEKGTGPIFIKEVSNADISMGDVATLSVTWIGIPKPKIQ 3489
 QY 59 ----- 58
 Db 3490 WFFNGVLLTPSADYKFVFDGDDHSLIILFTKLEDEGEYTCMASNDYGKTCISAYLKINSK 3549
 QY 59 ----- 58
 Db 3550 GEGHKDTETESAVAKSLEKLGPCPPHFLKLPICRCAOGLPAI FEYTVVGEPAPTVMF 3609
 QY 59 ----- 58
 Db 3610 KENKQLCTSVYTYI IHNPNGSGTFIVNDPQREDSGLYICKAENMLGESTCAAELLVLED 3669
 QY 59 ----- 58
 Db 3670 TDMTDPCKAKSTPEAPDPOTPLKGA VEAALDSEOEIATFVKDTILKAALITEENQOL 3729
 QY 59 ----- 58
 Db 3730 SYEHIKANELSQLPLGAQELQILQDKLTPESTREFLCINGSIHFPQPKPEPSNLQL 3789
 QY 59 ----- ILTEPEG- 65
 Db 3790 QIVOSQKTFSEKIGILMPEPETOAVLSDTEKIFPSAMSIEQINSLTVEPLKTLAEPEGN 3849
 QY 66 ----- 65

Db 3850 YPOSSIPEPPMHSYLTSAEEVLSLKEKTVSDTNREORVT LQKQBAQSALILSOSLAEGHV 3909
 QY 66 ----- 65
 Db 3910 ESLQSPDVMISQVNYEPLVPSEHSTEGGKILIESANPLENAGQDSAVRIEKGKSLRPL 3969
 QY 66 ----- 65
 Db 3970 ALEEKQVLLKEEHSNDVMPDPQIIIESKREPVAIKKQVQGRDLLSKESLLSGIPERQ 4029
 QY 66 ----- 65
 Db 4030 LNLKIQICRALQAAVASQPGLFSEWLRNIEKVEAVEAVNI TOEPRHIMCMVLTSAKSVT 4089
 QY 66 ----- 65
 Db 4090 BEVTIIIEDVDPMANLKMURDALCAIYBEIDILTAEGPRI QOGAKTSLQEEHDSFG 4149
 QY 66 ----- 65
 Db 4150 SOKVEPITEPEVESKYLISTEEVS YFNQSRVKYLDATPVTKGVASVVDKODESLKP 4209
 QY 66 ----- 65
 Db 4210 SEEKESSSESGTEEVATVKIOEAEGLIKEDGPMIHTPLV DTVSEGDIVHLTTSITNA 4269
 QY 66 ----- 65
 Db 4270 KEYNMYFENKLVPSDEKFKCLOQONTYTLVIDKNTEDH OGEVYVCEALNDSGKTATSXL 4329
 QY 66 ----- APNI----- 69
 Db 4330 TVVKRAAPVTKRIEPLVALGHLAKFTCEIO SAPNVRPQWPKAGREIYESDKCSIRSSK 4389
 QY 70 ----- 69
 Db 4390 YISSLEILRTQVVDGCEYTKASNEYGSVSTATLT VTPVGGEKKVKRLLPERKPEPKE 4449
 QY 70 ----- 69
 Db 4450 VLVKSLVRKRPESEEPKVPKLEKVKPAVPEPPPPKVP EEVEVPTVKRERKIPEPTK 4509
 QY 70 ----- 69
 Db 4510 VPEIKPAIPLPAPEPKPAEAVKTIKPPPEPEPTPIA APVTVPVVGKKAERKAPKEA 4569
 QY 70 ----- 69
 Db 4570 AKPGPIKGVPKTPSPIEARRKLRPGSGGKPPDEAP TYQLKAVPLKAVKEIKDIL 4629
 QY 70 ----- 69
 Db 4630 TESEFVGSSAIFECVLSFSTAITTMMKDGSNIRESP KHRFIADGDKRKLHIIDVQLSDAG 4689
 QY 70 ----- 69
 Db 4690 EYTCVLRNLNKEKTSKALVVEELPVRFVKTLEEVTV VKGQPLYLSCELNKERDVMRK 4749
 QY 70 ----- 69
 Db 4750 DGKIWEKPGRIVPGVIGLMRALTINDADDDAGTYT TVTENANNLESCSCVKVVEIRD 4809
 QY 70 ----- 69
 Db 4810 WLVPKPIRDOHVKPGTAIFACDIAKDTPNIKWFKGY DEIPAEPNDKTSILRDGNHLYUKI 4869
 QY 70 ----- 69
 Db 4870 KNAMPEDIAEYAVEIEGKRYPAKUTLGERVELLKP IEDVTIYEKESASPDALISEADIP 4929
 QY 70 ----- 69

Db	4930	QOMKLGELLRSPSTCEIKAEGRFLTLHKVLQDAGEVLYQALNAITTAITVKEIEL	4989
Qy	70	-----	69
Db	4990	DFAVPLKDVTPERQARFECVLTREANVWSKGPDIISKSDKFDIIADGKHILVINDS	5049
Qy	70	-----	69
Db	5050	QFDBGVYTAIEVEGKTSARLFTVGTIRLKFMSPLEDQTVKEGETATFVCELSHEKHMVW	5109
Qy	70	-----	69
Db	5110	FKNDAKLHSTRVLISSEKTHLEMKVETLDDISQIKAQVKELSSSTAQLKVLEADPYFT	5169
Qy	70	-----	69
Db	5170	VKLHDKTAVEKDEITLKCEVSKDVPKWFKOGEEIVPSPKYSIKADGLRRIKIKKADLK	5229
Qy	70	-----	69
Db	5230	DKGEVVCDCGTDKTKANVTVEARLIEVEKPLYGVEFVGETAHEFIELSEPDVHGQWKLK	5289
Qy	70	-----	69
Db	5290	GOPLTASPDCEIIEGKKHILILHNCQLGTMGEVSFOAANAKSAANLKVKELPLIFITPL	5349
Qy	70	-----	69
Db	5350	SDVKVEKDEAKFECEVSREPRTFRMLKGTBEITGDRFELIKDGTQHSMTKSAAFEDE	5409
Qy	70	-----	69
Db	5410	AKYMPAEDKHTSGKLIIEGIRLKLTPDKVTAKESAVFTVELSHDNIRVKNFKNDQ	5469
Qy	70	-----	69
Db	5470	RLHTTRSVSQDEGKTHSITFKDLSIDDTSQIRVEAMGMSSEAKLTVLEGDYPFTGKLQD	5529
Qy	70	-----	69
Db	5530	YTGVEKDEVILCEISKADAPKMFKDGEIKPSKNAVIKTDGKRMILILKALKSDIQ	5589
Qy	70	-----	69
Db	5590	YTCDCGTDKTSGLDIEDREIKLRPLHSVEMETETARFETEISEDDIHANWKLGKAL	5649
Qy	70	-----	69
Db	5650	LQTPDCEIKEEGKIHSVLHNCRLDQGGVDFQANVKSSAHLRVKPRVIGLLRPLKDV	5709
Qy	70	-----	69
Db	5710	VTAGETATFDCELSYEDIPVEWYLGKLEPSDKVVPVPRSEGVHTLTLRDVKLEDA	5769
Qy	70	-----	69
Db	5770	LTAKDFKTHANLKVKEPPVEFTKPLEDQTVKEGATAVECEVSRENKVKWFKNGTEILK	5829
Qy	70	-----	69
Db	5830	SKKYEIVADGRVKLVHDCTPEDIKTYTCDKDFKTSNCLNVPPHVEFLRPLTLQVR	5889
Qy	70	-----	69
Db	5890	EKEMARFECBSRENKVKWFKDGAIEIKGKYDIISKGAVRILVINKCLLDDDEA	5949
Qy	70	-----	69
Db	5950	VRTARTSMLTVLEEBAVFTKLANIEVSETDIKLVCYSKGAIEVIWYKGDDEEIIETG	6009
Qy	70	-----	69
Db	6010	RYEILTEGRKRILVIONAHLEDAGNYNCRLPSSRTDGVKVKVHELAAEFISKQNL	6069
Qy	70	-----	69
Db	6070	EKAFFVCSISKSEFPVQWKRRDDKTLESGBKYDVIADGKKRVLVVKDQATLQDMCT	6129
Qy	70	-----	69
Db	6130	AARAAHLTVIEKLRIVVPLKOTRVKEQOEVPNCEVNTGAKAKWFRNEAIFDSKYI	6189
Qy	70	-----	69
Db	6190	ILQKDLVYTLIRDAHLDDQANYNSLTNHRGENVKSAAANLIVEEEDLRIVEPLKDIETM	6249
Qy	70	-----	69
Db	6250	EKKSVTFCWKVNLNVLTKWTKNGEEVPFONRVSYRYDKYKHLMTIKDCGFPDGEYIVT	6309
Qy	70	-----	69
Db	6310	AGDKSVAELLIIIEAPTEFVEHLEDQTVTFDDAVFSCQSREKANVKYRNGREIKEGK	6369
Qy	70	-----	69
Db	6370	KYFEKDGSIHRLIIKDCRLDDECEYACGVEDRKSRLRFVEEIPVEIIRPPDILEAPG	6429
Qy	70	-----	69
Db	6430	ADVFLAELNKDKVEVQWLRNNMVVQGDKHQMSEGIHRLQICDIKPRDQGEYRFAIK	6489
Qy	70	-----	69
Db	6490	DKEARAKLELAAPKIKTADQDLVVDGKPLTMVVPYDAYPKABAEWFKENEPLSTKID	6549
Qy	70	-----	69
Db	6550	TTAOTSFRILEAKGDKGRYKIVLQNHGKABEFINLKVDPVGRNLEVTETPDGEV	6609
Qy	70	-----	69
Db	6610	SLAMEEPLTDGSKIIGVWVERRDIKRKTWVLATDRAESCEFTVTGLQKGVLYFRVSA	6669
Qy	70	-----	69
Db	6670	RNRVGTGEVETDNPVEARSKYDVGPPPLNVTITDVRNFGVSLTWEPPEYDGGAEITNV	6729
Qy	70	-----	69
Db	6730	IELRDKTSIRWDATVRAEDLSATVTDVWEGQEYSFRVRAQNRIGVGPASAATPFVKVA	6789
Qy	70	TANLTSS-----	76
Db	6790	DPIERPPVNLTSQTSQSSVQLKWPPLKDGSPILGYIIERCEGKNWIRCNMKLV	6849
Qy	77	-----	76
Db	6850	PELTYKVTGLEKGNKLYRVAENKAGVSDPSEILGPLTADDAFVEPTMDLSAFKDGLEV	6909
Qy	77	-----	76
Db	6910	IVPNPITILVPSTGYPRPTATMCFGDKVLETDGRVMKTL SAYAELVISPSERSDKIYT	6969
Qy	77	-----	76
Db	6970	LKLENRVKTSIGEIDVNVIAIARPSAPKELKFGDITKDSVHLTWEPDDDDGGPLTGYVVEK	7029
Qy	77	-----LLSVC-----	81
Db	7030	REVSRTKVMDFVTDLEFTVPDLVQSGKEYLKFVCARNKCGPGEPAYVDEPNMSTPAT	7089
Qy	82	-----	81
Db	7090	VPDPENVKWRDRTANSIFLTWDPKNDGGSRIKGYIVERCPRGSKWVACGEPVAETKM	7149

Qy	82	-----	81
Db	7150	EVTGLEEGKMYAVRVKTLNRQASKSPRTEIEIQAVDTQEAPEIFLDVVKLAGLTVKACT	7209
Qy	82	-----	81
Db	7210	KIELPATVTVGKPEPKITWTAKDMILKQDKRITIENTVPPKSTVTIVDSKESDGTGTYIEAV	7269
Qy	82	-----	81
Db	7270	NVCGRATAVVENVLDKPGPAAFDITDVTNESCLLTWNPPRDDGGSKITVYVVERRATD	7329
Qy	82	-----	81
Db	7330	SEVHKLSTVKOTNFKATKLIIPNKEYIFRVAENMYGAGEPVQASPIAKYQFDDPGGP	7389
Qy	82	-----	81
Db	7390	TRLEPSDITKDAVTLTWCEPDDGGSPITGYWVERLDPDTDKWRCNKMPVKDITTYRVKG	7449
Qy	82	-----	81
Db	7450	LTNKKYFRVRVAENLAGCGKSPKSTEPILIKDIPDPPMPGKPTVKDVGKTSVRLNWTK	7509
Qy	82	-----	81
Db	7510	PEHGGAKIESYVIEMLKCTGTDENVRVAEGVPTTOHLLPGLMEGOEYSFRVRAVNKAGES	7569
Qy	82	-----	81
Db	7570	EPSESPDVLCREKLYPPSPRWLEVINITKNTADLKWTVPEKDDGGSPITNIYIVEKRDVR	7629
Qy	82	-----	81
Db	7630	RKGHTVDTTKTKTCTVTLTEGLYVFRVAENAIGQSDYTEIEDSVLAKDTFTTPGP	7689
Qy	82	-----	81
Db	7690	PYALAVVDVTKRHVDLKWPEPKNDGGRPIORYVIEKKERLGRVWKAGTAGDCNFRVT	7749
Qy	82	-----	81
Db	7750	DVIEGTEVOFVRAENAGVCHPSEPTIELSIEDTSPSPPLDLHVTDAGRKHIAIAWK	7809
Qy	82	-----	81
Db	7810	PPEKNGSPIIIGYHVEMCPVTEKMRVNSRPIKDLKFVBEGVVPDKEYVLRVAVNAI	7869
Qy	82	-----	81
Db	7870	GVSEPSEISENVAKDPCKPTIDLETHDIIIVIEGKLSIPVPPRAVPVPTVSHKDGKE	7929
Qy	82	-----	81
Db	7930	VKASDLTMKNDHISAHLEVPKSVRADAGIYITITLENKLSATASINVKIGLPGPCKDI	7989
Qy	82	-----	81
Db	7990	KASDITKSSCKLTWEPPEFDGTPILHYVLERREAGRTYIPVMSGENKLSWTVKDLIPN	8049
Qy	82	-----	81
Db	8050	GEYFPRVAVNKVGGEYIELKNPVIADPKQPPDPVDVEVHNPTAEAMITWKPLPYD	8109
Qy	82	-----	81
Db	8110	CGSKIMGYIEKIAKEERWKRNEHLVPILTYTAKLEGKEYQFRVRAENAAAGISEPS	8169
Qy	82	-----	81
Db	8170	RATPPTKAVDPIDAPKVILRTSLEVKGDEIALDASISGSPYPTITWIKDENVIIVPEIK	8229
Qy	82	-----	81

Db	8230	KRAAPLVRRRKGEVQEBEPFVLPLTORLSIDNSKKGESQSLRVRDSLRPDHGLYMIKIVEND	8289
Qy	82	-----	81
Db	8290	HGIAKAPCTVSVLDTPOPPINFVPEDIRKTVLCKWBPPLDDGGSEIINYITLEKKDKTKP	8349
Qy	82	-----	81
Db	8350	DSEWIVVTSTLRHCKYSVTKLIEGKEYLFRVRAENRFGPGPCVSKPLVAKDPFGPPDAP	8409
Qy	82	-----	81
Db	8410	DKPIVEDVTSNMLVKWNEPKDNGSPILGYWLEKREXVNSTHWSRVNKSLLNALKANVDGL	8469
Qy	82	-----	81
Db	8470	LEGLTYFRVCAENAAAGKFSPPSDPKTAHDPIPSPOGPIPRVTDTSSTTIELEWEPPA	8529
Qy	82	-----	81
Db	8530	FNGGGEIVGYFVDKQLVGTNKNKSRCTEKMIKVROYTVKEIREGADYKLRVSAVNAAGEP	8589
Qy	82	-----	81
Db	8590	PGETQPTVABPQEPPEPAVELDVSVKGGTQIMAGKTLRIPAVVTCRPVPTKVMTKEEGELD	8649
Qy	82	-----	81
Db	8650	KORVVIDNVGTSKSELIIDALRKDHGRYVITATNCSGSKFAAARVEVFDVPGVLDLKPV	8709
Qy	82	-----	81
Db	8710	VTRNKKLLNMSDPEDDGGSEITGPIIBRKDAKMHTRQPIETERSKCDITGLLEGQBYK	8769
Qy	82	-----	81
Db	8770	FRVIAKNKFCGPPVEIGPILAVDPLGPPTSPELTYTERQSTITLQWKEPRSGSGPI	8829
Qy	82	-----	81
Db	8830	QGYIEKRHRKDPDFERNKRLCPTTSFLVENLDEHOMVEFRVKAIVEGESEPSLPLNV	8889
Qy	82	-----	81
Db	8890	VIQDDEVPTIKLRLSVRGDTIKVKAGEPVHIPADVTGLPMPKIEWSKNETVIEKPTDAL	8949
Qy	82	-----	81
Db	8950	QITKEVSRSEAKTELISIPKAVREDKGTVTVTASNRLGSVFRNVHVEVYDQSPPRNLAV	9009
Qy	82	-----	81
Db	9010	TDIAESCYLTWDAPLDNGGSEITHYVVDKRDASRKKAEWEVNTAVKRYGIMKLI PN	9069
Qy	82	-----	81
Db	9070	GQYEFVRVAVNKYGISDECKSDKVIQDPYRLPGPPGPKVLARTKGSMLVSWTPPLDNG	9129
Qy	82	-----	81
Db	9130	GSPITGYWLEKREEGSPYWSRVRAPITKVGLKGVEFNVPRLLLEGVKYQFRAMAINNAGI	9189
Qy	82	-----	81
Db	9190	GPFSESDPEVAGDPIFPPOPPSPCEVKDKTKSSIISLGMKPPAKDGGSPKIGYIVEMQEE	9249
Qy	84	-----	83
Db	9250	GTTDWKRNVNPDKLITTCCEVWPNLKLRYFRVKAIVEGESEPSDITTEIPATDIOE	9309
Qy	84	-----	83

Db 9310 EPEVFIDIGAQCCLVKAGSQIRIPAVIKGRPTPKSSWEFDGKAKKMGVHDIPEDAQ 9369
Qy 84 ----- 83
Db 9370 LETAENSSVIIPECKRSHTGYSITAKNKAQGTANCVRKVMVDPGPKDLKVSIDITRG 9429
Qy 84 ----- 83
Db 9430 SCRLSKWKPDDGGDRIKGVYIEKRTIDGKAWTKYNPCDGSFTFVVPDLLSEQQYFFVRV 9489
Qy 84 ----- 83
Db 9490 AENRFGIGPPVETIORTTARDPIYPPDPPIKLGILITKNTVHLSWKPPKNDGGSPTHY 9549
Qy 84 ----- 83
Db 9550 IVECLAWDPTGKKEAMROCNKRDEELOFTVEDLVEGGEYEFVRKANNAAGVSKPSATV 9609
Qy 84 ----- 83
Db 9610 GPDCORPMPSSIDLKEFMEVEEGTNVNI VAKIKGVPPTLTWFKAPPKPDNKEPVLV 9669
Qy 84 ----- 83
Db 9670 DTHVKNLVDDTCTLVIPOSRRSDTGLYITITAVNNLGTASKEMRLNVLGRPGPVGPIKF 9729
Qy 84 ----- 83
Db 9730 ESVAQDMTSLWPPPKDGGSKI TNVIEKREANRKTWVHVSSEPKECTYTITPKLEGHE 9789
Qy 84 ----- 88
Db 9790 YVFRIMAQNKYGIGELDSEPEARNLFSVPGAPDKPTVSSVTRNSMTVNWEEPEYDGS 9849
Qy 89 ----- 88
Db 9850 PVTGYWLEMKDTSKRWRVNRDPIKAMTLGVSYKVTGLIEGSDYQFRVYAINAAGVGA 9909
Qy 89 ----- 88
Db 9910 SLPSDPATARDPIAPGPPPKVTDWTKSADLEWSPPLKDGSKVTGYIYEYKEGKEE 9969
Qy 89 ----- 88
Db 9970 WEKGDKEVRGTLVVTGLKEGAFYKFRVSANVIAGIGEPGEVTDVIEMKDLVSPDLQL 10029
Qy 89 ----- 88
Db 10030 DASVRDRIVVHAGGVIRIIAYVSGKPPPTVTNMNERTLPQEATIIETAISSSMVKNCO 10089
Qy 89 ----- PEDD--- 92
Db 10090 RSHGVVYLLAKNEAGERKKTIIIVDVLDPGPGVGTPLAHLNLTNESCCLTWFSPEDDGGS 10149
Qy 93 ----- 92
Db 10150 PITNYVIEKRESORAWTPVTYVTRONATVQGLIQGAYFFRIAAENSIGMPFVETSE 10209
Qy 93 ----- 92
Db 10210 ALVIREPITVPERPEDLEVKEVTNVTLTWNPYPKYDGGSEIINYVLESRLICTEKFHKV 10269
Qy 93 ----- 92
Db 10270 TNDNLSRKVTYKLGKEGDTVEYRVSAVINVGQKPSFCTKPTCKDELAPPTLHLDFRD 10329
Qy 93 ----- 92
Db 10330 KLTIIRGEAFALTGRYSKPKPKVSWFKDEADVLEDDRTHIKTTPATLALEKIKAKRSDS 10389
Qy 93 ----- 92
Db 10390 GKVCVVENSTGRKGFQCNVVDHPGPPVGPVSFDEVTKDYMVISMKPPPLDDGGSKI TN 10449

Qy 93 ----- 92
Db 10450 YIIEKVEKGVDMVMPVTSASAKTTCKVSKLEGGDYIFRIHAENLYGISDPLVSDSMKAK 10509
Qy 93 ----- 92
Db 10510 DRFRVPADPOPIVTEVTKDSALVTWNKPHDGGKPIITNYILEKRETMKRWARVTKDPIH 10569
Qy 93 ----- 92
Db 10570 PYTKFRVPDLLLEGQCYBFRVSAENEIGIGDPSBPKVFAKPIAKPSPVPVPEAIDTTC 10629
Qy 93 ----- 92
Db 10630 NSVDLTWQPRHDGSKILGYIVSYQVGDSEWRANHTPESCPETKYKVTGLRDGQTYK 10689
Qy 93 ----- TDPGH----- 97
Db 10690 FRVLAVNAAGESDPAHVPEPVLVKDRLEPPPELILDANWAEQHIKVGDTLRLSAILKGVP 10749
Qy 98 ----- 97
Db 10750 FPKVTWKEDRDAPTKARIDVTPVGSKLEIRNAAHEDGGIYSLTVENPAGSKTVSVKVLV 10809
Qy 98 ----- 97
Db 10810 LDKPGRDLEVBIRKDSCLYLTWKEPLDDGGSVITNVVERRDVAQAQWSPLSATSKK 10869
Qy 98 ----- 97
Db 10870 SHFAHLNEGQYLFRVAENQYGRGPFVETPKPIKALDPLHPGPKDLHHVDVDKTEV 10929
Qy 98 ----- 97
Db 10930 SLVWNKPRDGGSPITGYLVEYQSEGTQDMIKFKVTNLECVVTLQOGKTYRFRVKAEN 10989
Qy 98 ----- 97
Db 10990 IVGLPLPTTPIECQEKLVPPSVVELDVKLEGLVVKAGTTVRFPPIIRGVVPVPTAKWTT 11049
Qy 98 ----- 97
Db 11050 DGSEIKTDEHYTVETDNFSSVLTIKNCLRRDTGEOITVSNAGSKTVAVHLTVLDVPGP 11109
Qy 98 ----- 97
Db 11110 PTGPIINILDVTPHEMTISWQPKDDGGSPVINIIVEKODTRKDTMGVVSSGSKTKLKIP 11169
Qy 98 ----- 97
Db 11170 HLQKGEYVFRVRAENKIGVGPLDSTPTVAKHKFSPSPGPKPVVTDITENAAATVSWTL 11229
Qy 98 ----- 97
Db 11230 PKSDGSPITGYMERREVTGKWRVKNKPTIADLKFRVTGLYEGNTYEFVRFAENLAGUS 11289
Qy 98 ----- ADLV----- 101
Db 11290 KPSPSDDPIKACRPKIPKPPGPPINPKLKDKSRETADLVWTKPLSDGGSPILGYVVECOKPG 11349
Qy 102 ----- 101
Db 11350 TAQWNRINKDELIROCAFVPGVPLIEGNEYRFRIKAANIVGEGEPRELAESVIAKDLHHP 11409
Qy 102 ----- 101
Db 11410 EVELDVTCDVTVRVGQTIRILARVKGRPEPDIITWTKEGKVLVREKRVLDLIQDLPRVEL 11469
Qy 102 ----- 101
Db 11470 QIKEAVRADHGKVIISAKNSSGHAQSAI NVNLDPRGPCONLKVNTVTKENCTI SWENPL 11529

QY	102	-----	101
Db	11530	DNGSEITNFIVEYRKPQKQMSIVASDVTKRLIKANLANNEYFRVCAENKVGQPTI	11589
QY	102	-----	101
Db	11590	ETKTPILAINPIDRGPENLHIADKGTFFYLKWRRPDYDGGSPNLSYHVERRLKGSD	11649
QY	102	-----	101
Db	11650	WERVHGSIKETHYMDRCVENQIYEFRTVOTKNEGSEDMVKTBEVVVKEDLOKPVLDL	11709
QY	102	-----	101
Db	11710	LSGVLTVKAGDTIRLEAGVRGQPEVAWTKDXDATLSTRPRVKIDTRADSSKFSLT	11769
QY	102	-----	101
Db	11770	KBSDGGKYVATNTAGSFVAYATVNVLDKPGVRLNKIVDVSSDRCTVCWDPEDDGG	11829
QY	102	-----	101
Db	11830	EIQNYILEKCEKRMVWSTYSATVLTGTTVTRLIEGNEYIFRVAENKIGTGPPTES	11889
QY	102	-----	101
Db	11890	VIAKTKYDKPGRPPPEVTKVSKEMTVMNPPEDYDGGKSITGYFLEKKEKHSTRV	11949
QY	102	-----	101
Db	11950	KSAIPERBMKVONLLPDHEYQFRVKAENEIGIGEPSLPSRPVAKDPIEPGPTNFR	12009
QY	102	-----	101
Db	12010	DTTKHSITLWGKPVYDGGAPIGYVEMRPKIADASPEGWKRCNAAQLVRKEFT	12069
QY	102	-----	101
Db	12070	LDENOYEFVCAQNVGIGRPAELKEAIKPEILEPPEILDASMRKLIVRAGCP	12129
QY	102	-----	101
Db	12130	FAIVGRPAPKVTRKVGIDNVVRKGOVDLVDTMFLVIPNSTRDDSGKSLTLVNP	12189
QY	102	-----	101
Db	12190	KAVFVNRVLDTPGPVSLKVSVDTKTSCHVSWAPPENDGGSQVTHYIVEKRE	12249
QY	102	-----	101
Db	12250	TVTPEVKKTSFHVNLVPGNEYFRVTAVNEYGPGVPTDVPKPVLASDPLSE	12309
QY	102	-----	101
Db	12310	ATEMTKNSATLAWLPLRDGGAKIDGYIISYREEEQPADRWTEYSWKDLSLV	12369
QY	102	-----	101
Db	12370	KKYKFRVAARNAVGVSLPREAGVYEAKQLLPPKILMPEQITIKAGKRLIEA	12429
QY	102	-----	101
Db	12430	HPTCKWKGEDEVTTSSHLAVHKADSSSILIIKDVTRKDSGYSLTAENSSG	12489
QY	102	-----	101
Db	12490	VWMDAPGPQPPFDISDIDACSLSWHIPLEDDGGSNITNYIVEKCDVSRG	12549
QY	102	-----	101
Db	12550	TKTSCRVGKLIPIQOEYIFRVAENRFGISEPLTSPKMVAQFPFVPEPKNAR	12609
QY	102	-----	101

Db	12610	CIFVADRDPDGGSPIIIGYLIERKERNLLWVKANDTLVRSTEYPCAGLVEGLE	12669
QY	102	-----	101
Db	12670	YALNAGSPPSPKPTTEYVYARMFVDPGKPEVIDVTKVSLIWARPKHDKGSKI	12729
QY	102	-----	101
Db	12730	EACKLPDQKWRCTAPHQIQOEYATATCLEEKAQYQFRAIARTAVNISPPSEP	12789
QY	102	-----	101
Db	12790	LAENVPRIIDLSVAMKSLTLVKAGTNVCLDATVFGKPMPTVSWKDKGTLKPA	12849
QY	102	-----	101
Db	12850	QRNLCTLEFVNRKDSGDYTTAENSSGSKSATIKLVLDKPGPPASVKINKMY	12909
QY	102	-----	101
Db	12910	LSWEPLEDGGSEITNYIVDKRETSRPNQAVSATVPITSCSVEKLIEGHEY	12969
QY	102	-----	101
Db	12970	KYGVGDPVFTEPAIAKNPYDPPGRCDPPVISNITTKDHTVSWKPPADGG	13029
QY	102	-----	101
Db	13030	RETQAVNWKVARKPIERTLKATGLOBOTEYEFRTAINKAGCPKPSDASKA	13089
QY	102	-----	101
Db	13090	YPPAPPAPFKYVDTTRSSVSLSGKPAYDGGSPIIIGYLVEVKRADS	13149
QY	102	-----	101
Db	13150	TRPEVTGLMEDTOYQPRVAVANKIGYSDPSDVPDKHYPKDILIPPEGE	13209
QY	102	-----	101
Db	13210	AGVTMRLYVPVKGRRPPKKTWSKPNVNLDRIGLDIKSTDFTFLRCEN	13269
QY	102	-----	101
Db	13270	TLENSCGKKEYTIIVKVLDTPGPPINVTVKEISKDSAYVTWBPPIIDG	13329
QY	102	-----	101
Db	13330	DAERKSWSTVTTECSKTSFRPNLEEGKSYFPRVFAENEYIGDPGETR	13389
QY	102	-----	101
Db	13390	VVDLKVRVSKSSCSIGWKPHSDGSRIGYVVDVTEENKQWRVMKSLSL	13449
QY	102	-----	101
Db	13450	EGKEYTFRVSAENENGEGTPEITVVARDDVAPDLKGLPDLCLAKENS	13509
QY	102	-----	101
Db	13510	KGKPAFVSWKKGEDPLATDTRVSVESAVNTTLIVYDCKSDAGKYTITL	13569
QY	102	-----	101
Db	13570	TISIKVKGPGIPTGPIKPFDEVTAEAMTLKWAAPPKDDGGSEITNYIL	13629
QY	102	-----	101
Db	13630	ASAVQKTTFRVTRLHEGMEYTFRVAENKYGVGEGKSEPIVARHPDFD	13689
QY	102	-----	101

Db	13690	VRHDSVSLTWDPKKTTGGSPITGVHLEPKERNLSLLWKRANKTPIRMDFKVTGLTEGLE	13749
Qy	102	-----	101
Db	13750	EFRVMAINLAGVGRSLPSEPVVALDPIDPPGKPEVINITRNSVTLIWTEPKYDGGHKL	13809
Qy	102	-----	101
Db	13810	GYIVEKRDLPKSKWMKANHVNPECAFTVTDLVEGGKVEFRIRAKNTAGALSAPSESTET	13869
Qy	102	-----	101
Db	13870	IICKDEYEAPTIVLDPTIKDGLTIKAGDTIVLNAISILGKPLPKSSWSKAGDIRPSDIT	13929
Qy	102	-----	101
Db	13930	QITSTPTSSMLTIKYATRDKAGEYTTATNPFGTKEVHKVTVLDVPGPPGVSEISNVA	13989
Qy	102	-----	101
Db	13990	EKATLTWTPPLEDGGSPIKSYILEKRETSRLLWTVVSEDIQSCRHVATKLIQGNIEYIFRV	14049
Qy	102	-----	101
Db	14050	SAVNHYGKEPVQSEPVKMDRFGPPGPEKPEVSNVTATVSWKRPVDDGGSEITGY	14109
Qy	102	-----	101
Db	14110	HVERREKSLRWRAIKTPVSDLRCKVTGLOEGSTYEFVRSAENRAGIGPPSEASDSVLM	14169
Qy	102	-----	101
Db	14170	KDAAYPPGPPSNPHVTDITTKSASLAWGKPHYDGGLEITGYVVEHQVDEAWIKDITGT	14229
Qy	102	-----	107
Db	14230	ALRITQFVVPDLQTEKYNFRISAINDAGVGEPAVDPVEIVEREMAPDFELDAELRRTL	14289
Qy	108	-----	107
Db	14290	VVRAGLSIRIFVPIKGRPAPEVTWKDNLNLKNRANIENTESFTLLIIPECNRYDTGKFV	14349
Qy	108	-----	107
Db	14350	MTIENPAGKSGFVNVRLDTPQVNLRLRPTDITKDSVLHWDPLIDGGSRTITNVIVEK	14409
Qy	108	-----	107
Db	14410	REATKRSYSTATTCKCHKCTYKVTGLSEGCEYFVRVMAENEYIGIGEPTETTEPVKASEAPS	14469
Qy	108	-----	107
Db	14470	PPDSLNMIDITKSTVSLAWPKPKHGGSKIITGVIEAQRKGSQDQTHITTVKGLECVRN	14529
Qy	108	-----	107
Db	14530	LTEGEYTFQVMAVNSAGRSAPRESRPVIVKEQTMPELDLRCIYOKLVIKAGDNKIVE	14589
Qy	108	-----	107
Db	14590	IPVLRPKPTVWKKGQDILKQTVNFETTATSTILNINECVRSDSGPYPLTARNIVE	14649
Qy	108	-----	107
Db	14650	VGDVITIQVHDIQCPPTGPIKPFDEVSDFVTFSDPPENDGGVPISNYVVEMQTDSITW	14709
Qy	108	-----	107
Db	14710	VELATTVIRTYKATRLTTGLEQFRVKAQNRVGVGPGITSAMIVANYPKVPGPPTQ	14769
Qy	108	-----	107
Db	14770	VTAVTKDSMTISWHEPLSDGSGPILGYHVERKERNILWQTVSKALVPGNIFKSSGLTDG	14829
Qy	108	-----	107
Qy	108	-----	107
Db	14830	IAYEPRVIAENMAGSKSPKSEPLALDPIIDPPGKPLNITRHTVTLUKAKPEYTGGF	14889
Qy	108	-----	107
Db	14890	KITSYIVEKRDLPNGRWLKNFNSNILENEFTVSLGTEDAAEFVRVIAKNAAGAISPPSEP	14949
Qy	108	-----	107
Db	14950	SDAITCDDVEAPKIKVDVKFDVTILKAGEAPRLEADVSGRPPPTMWSKDGELEGT	15009
Qy	108	-----	107
Db	15010	KLEIKIADFSTNLVKNKSTRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPPPEGLAVTE	15069
Qy	108	-----	107
Db	15070	VTSEKCVLSWFPPLDDGGAKIDHYIVQKRETSRLANTNVASEVQVTKLVTKLLKGNEYI	15129
Qy	108	-----	107
Db	15130	FRVMAVNVKVGGELESEPVLA VNPYGPDPKPNPEVTITKDSMVVCMGHGPDSDGSEI	15189
Qy	108	-----	107
Db	15190	INVIVERRDKAGORWIKCNKKTLDLRYKVGSLTEGHEYEFRIEMAENAGISAPSTSPF	15249
Qy	108	-----	107
Db	15250	YKACDVTFKPGPPGNPRVLDTSRSSISIAWNKPIYDGSSEITGYMVEIALPEDEMQIVT	15309
Qy	108	-----	107
Db	15310	PPAGLKATSYITGLTENQEKIRIYAMNSEGLGEPALVPGTPKABDRMLPPEIELDADL	15369
Qy	108	-----	107
Db	15370	RKVVTIRACCTLRFLVPIKGRPDPEVKWARDHGESLDKASIESASSYTLILVGNVRPDS	15429
Qy	108	-----	107
Db	15430	GKYILTVENSSGSKSAFNVVRVLDTPGPPQDLKVKEVTKTSVTLTWDPPLLDGGSKIKNY	15489
Qy	108	-----	107
Db	15490	IVEKRESTRKAYSTVATNCHKTSMKVQLQEGCSYFRVLAENEYIGLPAETAESVKAS	15549
Qy	108	-----	107
Db	15550	ERPLPPGKITLMDVTRNSVLSWEKPEHGGSRILGVIVEMQTKGSDKWATCATVKVTEA	15609
Qy	108	-----	107
Db	15610	TITGLIOGEYSFRVSAQNEKGISDPRQLSVPIAKOLVIPPAPKLLPNTFTVLAGEDLK	15669
Qy	108	-----	107
Db	15670	VDVPFGRPTPAVTHWKDNVPLKQTRVNAESTENNSLLTIKDACREDVGHVYVVKLTNSA	15729
Qy	108	-----	107
Db	15730	GEAIELNVILDKPGPTGPVKMDEVTAISITLSWGPYPKYDGGSSINNYIVEKRDSTTT	15789
Qy	108	-----	107
Db	15790	TWQIVSATVARTTIKACBLKTGCEYQFRIAAENRYGKSTYLNSEPTVAQYFPKVPGPPTQ	15849
Qy	108	-----	107
Db	15850	PVVTLSSRDSMEVQWNEPISDGGSRVIGYHLERKERNILWVKNKTPIPOTKFKTTGLE	15909
Qy	108	-----	107

Qy 108 ----- 107
Db 15910 EGVEYEFVSAENIVGIGKPSKVSECVYVARDPCDPGRPEAIIIVTRNSVTLOWKKPTYDG 15969
Qy 108 ----- 107
Db 15970 GSKITGYIVEKKELPEGRMKASFTNIIIDHFEVTLGVEDHRYEFVRVARNAGVSEPS 16029
Qy 108 ----- 107
Db 16030 ESTGAITARDEVDPRISSMDPKYKDITIVHAGESFKVDADIYKPIPTIQWIKDQELSN 16089
Qy 108 ----- 107
Db 16090 TARLEIKSTDATSLSVKDAVRDYSNGNYILKAKNVAGERSVTNVKVLDRPGPEGVVI 16149
Qy 108 ----- 107
Db 16150 SGVTAECTLANPPLQDGGSDIINYIVERRETSRLVWTVVDANVQTLSCVKTKLEGNE 16209
Qy 108 ----- 107
Db 16210 YTFRIMAVNKYGVGEPLSEBPVAKNPVVDPAPKAPEVTTVTKDSMIVVWERPASDGG 16269
Qy 108 ----- 107
Db 16270 EILGVLEKRDKEGIRWTRCHKRLIGELRLRVLTGLIENHDEYEFVSAENAGLSEPPS 16329
Qy 108 ----- 107
Db 16330 AYQKADPIYKPGPPNPKVIDITRSSVFLSWKPIYDGCCEIOGYIVEKCDVNVGEWTM 16389
Qy 108 ----- DL 109
Db 16390 CTPTGINKTNI EVEKLEKHEYNFRICAINKAGVGEHADVPSPIIVEEKLEAPDIDL 16449
Qy 110 EL ----- 111
Db 16450 ELRKIINIRAGGSURLFVPIKGRPTPEVKMGKVDGEIRDAIIDVTSSTSLVDNRY 16509
Qy 112 ----- 111
Db 16510 DSGKYLTLNSSGTSKSAFVTVRLDTPSPPVNLKVTEITKDSVITWEPPLDGGSKIK 16569
Qy 112 ----- 111
Db 16570 NYIVEKREATRKSAAVVTNCHKNWKIDQLQEGCSYFVRVTAENEYIGLPAQTADPIK 16629
Qy 112 ----- 111
Db 16630 VAEVPPPGKITVDDVTRNSVLSWTKPEHDGSKIIQYIVEMQAKHSEKSECARVKSL 16689
Qy 112 ----- 111
Db 16690 QAVITNLQGEYLFVRVAVNEKGRSDPRSLAVPIVAKDLVIEPDVKPAFSSYSVQVQD 16749
Qy 112 ----- 111
Db 16750 LKMEVPISGRPKPTITWKDGLPLKQTTTRINVTDSLDLTLSIKETHKDDGGQYGITVAN 16809
Qy 112 ----- 111
Db 16810 VVGOKTASIEIVTLDKDPKPGPVKFDVDSAESITLSWNPPLYTGGCOITNIIYVQKRD 16869
Qy 112 ----- 111
Db 16870 TTWVDVVSATVARTTLKVTCLKTGTEYQFRIFAENRYGQSFALSDPIVAQYYPKEGPP 16929
Qy 112 ----- 111
Db 16930 GTPFAAISKDSMVQIOWHEPVNNGGSPVIGYHLERKERNILWTKVNTIIHDQFKAQN 16989
Qy 112 ----- 111

Db 16990 LEEGIEYEFVYAENIVGVGKASKNSECVYVARDPCDPTPEPIMVKRNEITLOWTKPVY 17049
Qy 112 ----- 111
Db 17050 DGGSMITGYIVEKRDLPDGRMKASFTNVIETQFTVSGLTEDORYEFVRIAKNAAGAIK 17109
Qy 112 ----- 111
Db 17110 PSDSTGPITAKDEVELPRISMDPKFRDTIVVNAGETFRLEADVHGKPLPTIEMLRGDK 17169
Qy 112 ----- PGNR 116
Db 17170 ESARCEIKNTDFKALLIVKDAIRIDGGQYILRASNVAGSKSPVNVKVLDRPGPEGPV 17229
Qy 117 QVRGVT ----- OLGG ----- 126
Db 17230 QVTGVTSEKSLTWPPLQDGGSDISHYVVEKRETSRLAWTVASEVVTNSLKVTKLLEG 17289
Qy 127 ----- 126
Db 17290 NEYVFRIMAVNKYGVGEPLSEAPVLMKNPFVLPGPPKSLEVTNIAKDSMTVCWNRPSDG 17349
Qy 127 ----- 126
Db 17350 GSEIIGYIVEKRDGRGIRWIKCNKRITDLRLRVLTGLTEDHEYEFVSAENAGVGEPS 17409
Qy 127 ----- ACSP ----- 130
Db 17410 ATVYKACDPVFKPQPPPTNAHIVDTTKNSITLAWCKPIYDGGSEILGYVVEICKADEEW 17469
Qy 131 ----- 130
Db 17470 QIVTQTLGRLVTRFISKLTEHQYKIRVCALNKVGLGEATSVPGTVKPEDKLEAPDL 17529
Qy 131 ----- TWS ----- 133
Db 17530 DSELARKGIWVAGGSARITHIPKGRPMPEITWSREEGEFTDKVQIEKGVNYTQLSIDCD 17589
Qy 134 ----- 133
Db 17590 RNDAGKYLKLENSGSKSAFVTVKVLDPGPPQNLAVKEVRKDSAFVWEPP1IDGGAK 17649
Qy 134 ----- 133
Db 17650 VKNYVIDKRESTRKAYANVSSKCSKTSFKVENLTEGAIYFVRMAENEFVGVPVETVDA 17709
Qy 134 ----- 133
Db 17710 VKAAEPSPPGKVTLTVDVTSQTSASLWMEKPEHDGGSRLGVYVEMQPKTKMSIVAESK 17769
Qy 134 ----- 133
Db 17770 VCNVVTGLSSGOEYQFRVKAYNEKKS DPRVLGVPIAKDLTIQPSLKL PNTYSIQAG 17829
Qy 134 ----- 133
Db 17830 EDLKIEIPVIGRPRPNI SWVKDGEPLKQTTVRNVEETATSTVLHIKEGNKDDFGKYTVTA 17889
Qy 134 ----- 133
Db 17890 TNSAGTATENLSVIVLEKPGPPVPRFVFEVSADFFVISWEPYPAYTGGCOISNYIVEKRD 17949
Qy 134 ----- 133
Db 17950 TTTTTHMVSATVARTTIKITKLKTGTEYQFRIFAENRYGKSAPLDSKAVIVQYPFKEPG 18009
Qy 134 ----- 133
Db 18010 PPGTPTVTSISKDMLVQWHEPVNDGGTKIIGYHLEQEKERNILWKLNTPIQDTKFKT 18069
Qy 134 ----- 133

Db 18070 TGLDEGLEVEFKVSAENIVIGICKSVSECFVARDPCOPGRPEAIVITRNNVTLKWKPK 18129
Qy 134 ----- 133
Db 18130 AYDGGSKITYIVEKKDLPGRMKASFVNLETFETVSGLVEDORYEPRVIARNAAGNF 18189
Qy 134 ----- 133
Db 18190 SEPSDSSGAITARDEIDAPNASLDPKYKDVIVVHAGETVLEADIRGKPIPDVWMSKDK 18249
Qy 134 ----- 133
Db 18250 ELEETAARMEIKSTIOKTLVAVKDCIRTDGOVILKLSNVGKTSIPITVKVLDKRGSP 18309
Qy 134 ----- 133
Db 18310 GPLKVTGVTAEKCYLAWNPPLODGGANISHYIIIEKRETSRLSWTQVSTEVALNKVKYL 18369
Qy 134 ----- 133
Db 18370 LPGNEVIFRMAVNKYGIGEPLESPVTACNPVKPPGPPSTPEVSAITKDSMVVTWARPV 18429
Qy 134 ----- 133
Db 18430 DDGTEIEGYILEKRDKEGVRWTKCNKKTLDLRLRVTLGLTEGHSYEFRAAENAAGVE 18489
Qy 134 ----- 133
Db 18490 PSEPSVFRACDALYPPGPPSNPKVTDTSRSSVSLAWSKPIYDGGAPVKGYVVEVKEAAA 18549
Qy 134 ----- 133
Db 18550 DEWTTCTPTGLOGKQFTVTKLKENTEYNFRICAINSEGVEPATLPGSVVAQERIEPPE 18609
Qy 134 ----- 133
Db 18610 IELDADLRKVVVLASATLRLFTVIKGRPEPEVKEAEGILTDRAQIEVTSFTMLVID 18669
Qy 134 ----- 133
Db 18670 NVTRFDSGRYNLTLENNSGKTAFAVNVRLDPSAPVNLTIREVKKDSVTLSEPPPLIDG 18729
Qy 134 ----- 133
Db 18730 GAKITNIVIEKRETRKAYATIINNCTKTFRIENLQEGCSYYFRVLASNEYGIGLPAET 18789
Qy 134 ----- 133
Db 18790 TEPVKVSEPLPGRVTLVDVTRNTATIKWKPESDGGSKITGYVVMQTKGSEKWSCT 18849
Qy 134 ----- 133
Db 18850 QVKTLEATISGLTAGEEYVFRVAANKEGRSDPRLQGVPIARDIEIKPSVELPFTFNV 18909
Qy 134 ----- 133
Db 18910 KAREQLKIDVPFKRPOATVNWKDGOTLKETTRVNVSSKTVTSLSIKEASKEDVGTYE 18969
Qy 134 ----- 133
Db 18970 LCVNSAGSITPITIIVLDRPGPPGPIRIDEVSCDSITISWNPPEYDGGCQISNIVEK 19029
Qy 134 ----- 133
Db 19030 KETTSTTHWIVSQAVARTSIKIVRLTTGSEYQFRVCAENRYGKSSYSESSAVVAEYFPSP 19089
Qy 134 ----- 133
Db 19090 PGPCTPKVHATKSTMLVTWQVPVNDGSRVIGYHLEYKERSILWSKANKILIADTQV 19149
Qy 134 ----- 133
Db 19150 KVSGLDEGLMEYRVAENIAGICKSKCEPVAPDCDPGQPEVTNITRKSLSLWS 19209

Qy 134 ----- 133
Db 19210 KPHYDGGAKITGIVERRELPGDGRWLKCNVTNIQETYFEVTELTEDORYEPRVFARNAAD 19269
Qy 134 ----- 133
Db 19270 SVSEPSSESTGPIIVKDDVEPRVMDVKPRDVIIVVAKAGEVLKINADIAGRPLPVISWAKD 19329
Qy 134 ----- 133
Db 19330 GIEIEERARTEIISTDNHNTLLTVKDCIRRDGTQVVLTKNVAGTRSVAVNCKVLDKPGPP 19389
Qy 134 ----- 133
Db 19390 AGPLEINGLTAEKCSLSWGRPQEDGGADIDYYHRRKKRSHLAWTICEGELQMTSCVKTK 19449
Qy 134 ----- 133
Db 19450 LLKNEYIFRVTGVNKGVGEPLESAIKALDPFTVPSPTSLEITSVTKESMTLCHSRP 19509
Qy 134 ----- 133
Db 19510 ESDGSEISGYIIERREKNSLRVVRVNVKPVYDLRVKSTGLREGCEYEYRVAENAAAGLS 19569
Qy 134 ----- 133
Db 19570 LPSETPLIRAEDPVFLPSPSPKPIVDSGKTTITIAWVKPLPDGGAPITGTYVEYKXSD 19629
Qy 134 ----- 133
Db 19630 DTDWKTSIOSLRGTEYITISGLTTGAEYVFRVKS NVKVGSDPSDSSDPOIAKEREPEPLF 19689
Qy 134 ----- 133
Db 19690 DIDSEMRKTLIVKAGASFTMTVPFRGRVPNVLWSKPDOLTRAYVDTTDSRTSLTIEN 19749
Qy 134 ----- 133
Db 19750 ANRNDGSKYTLTIQNVLSAASLTLLVVKVLDTPGPPTNITVQDVTKESAVLSWDVPENDGG 19809
Qy 134 ----- 133
Db 19810 APVKYHIEKREASKAWVSVTNNCNRLSYKVTNLOEGAIYYFRVSGENEFVGIGIPAEK 19869
Qy 134 ----- 133
Db 19870 EGVKITEKPSPEKLGVTISKDSVSLTWLKPHEHDGSRIVHYVVEALEKGQKNWVKCAV 19929
Qy 134 ----- 133
Db 19930 AKSTHHVVSGURENSEYFPRVFAENQAGLSDPRELLLPVLKEQLEPPEIDMKNFPSHTV 19989
Qy 134 ----- 133
Db 19990 YVRAGSNLKVLDIPISGKPLPKVTLSDRGVPLKATMRFNTEITAENLTINLKESVTADAGR 20049
Qy 134 ----- 133
Db 20050 YEITAANSSGTTKAFINIVLDRPGPTGPVVISDITEESVTLKWEPPKYDGGSOVTNYI 20109
Qy 134 ----- 133
Db 20110 LLKRETSTAVMTEVSATVARTMMKVMKLTGTGEYQFRIKAENRFGISDHDSACVTVKLP 20169
Qy 134 ----- 133
Db 20170 YTTGPPSTPWNTVNTRESITVGMHEPVNSGSAVVGYHLEMKDRNSILWQKANKLVIRT 20229
Qy 134 ----- 133
Db 20230 THPKVTIISAGLIYEFVRVAENAAAGVKPSHPSEPLAIDACEPPRNVRITDISKNSVSL 20289

Qy 134 ----- 133
Db 20290 SMOQAPDGGSKITGIYIVERDLDPGRWTKASFNTVETOFTISGLTONSQYSEFRVARN 20349
Qy 134 ----- 133
Db 20350 AVGSISNPSEVVGPIITCIDSYGGPVIDLPLEYTEVVKYRAGTSVKLRAGISGRPAITIEW 20409
Qy 134 ----- 133
Db 20410 YKODKEIQNALVCVENTDLASILKADRLNSGCYELKLRNAMASASATIRVOILDKP 20469
Qy 134 -----CLIT 137
|:|
Db 20470 GPPGGPIEFKVTVAEKITLLWRQPADDGGAKITHYIVREKRETSRVVMSVSEHLEECIIT 20529
Qy 138 ----- 137
Db 20530 TTKIKNNEYIFRVAVNKYIGIEPLESDSVVAKNAFVTPGPGIPEVTKITKNSMTVMW 20589
Qy 138 ----- 137
Db 20590 SRPIADGSDISGYFLEKRDKSLGFWKLVKETIRDRQKVTGLTENSQYQYRCAVNAA 20649
Qy 138 ----- 137
Db 20650 GQGPSEPFYKAADPIDPPGPAKIRIADSTKSSITLWGSKPVDGGSVAVTGYVVEIR 20709
Qy 138 ----- 137
Db 20710 QGBEEETVSTKGEVRTTEYVVSNLKPGVNYIFRVSAVNCAGOGEPIDENPEVQAKDIL 20769
Qy 138 ----- 137
Db 20770 EAPRIDLVALRTSIAKAGEDVQVLIPFKGRPPPTVTRKDEKNLGDARYSIENTDSS 20829
Qy 138 ----- 137
Db 20830 SLLTIPOVTRNDTGKYLITIENGVEPKSSTVSVKVLDTPAACQKLQKHVSRGTVLLW 20889
Qy 138 ----- 137
Db 20890 DPPLIDGSGPIINYIEKRDATKRTWSVSHKCSSTSPKLDLSEKTPFFFRVLAENEIG 20949
Qy 138 ----- 137
Db 20950 ICEPCETTEPVKAAEVPAPIRDLMSKSTKTSVILSWTKPDPDGGSVITEYVVERKKGKE 21009
Qy 138 ----- 137
Db 21010 QTMSHAGISKTCIEIVSQLKEQSULEFRVFAKNEKGLSDPVTIGPITVKELIITPEVDLS 21069
Qy 138 ----- 137
Db 21070 DIPGAQVTVRIGHNVHLELPYKGPKPSISMLKDGLPLKESEFVRFSKTENKITLSIKNA 21129
Qy 138 ----- 137
Db 21130 KKEHGKVTVIDNAVCAPIVITITLGPSPKPKPIRDEIKADSVILSDVPEDNGG 21189
Qy 138 ----- 137
Db 21190 GEITCYSIEKRETSQTNWKCSSVARTTFKVPNLVKDAEYQFVRAENRYCVSQPLVSS 21249
Qy 138 ----- 137
Db 21250 IIVAKHQFRIPGPPGKPIYINVTSDGMSLTWDAPVYDGGSEVTGFHVEKKERNLSILWQKV 21309
Qy 138 ----- 137
Db 21310 NTSPIGREYRATGLVEGLDYQFVVAENSAGLSPEPSKFTLAVSPVDPGTPDYIDV 21369
Qy 138 ----- 137

Db 21370 TRETTILKWNPLRDGSGKIVGYSIEKROGNERWRCNFTDVSECOYTVTGLSPCDRYEF 21429
Qy 138 ----- 137
Db 21430 RIIARNAVGTISPPSQSSGIIMTRDENVPPIVEFGEYFDGLIIKSGESLIKALVOORP 21489
Qy 138 ----- 137
Db 21490 VPRVTWFKDGVIEIKRMMEITNVLGSTSLFVRDATDRHRCVYTVVEAKNAGSKAKAEIKV 21549
Qy 138 ----- 137
Db 21550 KVQDTPGKVGPIRFTNITGEKMTLWMDAPLNDGCAPITHYIIEKRETSRLAWALIEDKC 21609
Qy 138 ----- 137
Db 21610 EAQSYTAIKLINGNEYQFRVSAVNFVGRPLDSDPVVAQIOYTVDPDAPGIPESNITGN 21669
Qy 138 ----- 137
Db 21670 SITLTWARPESDGGSEIQOYILERREKSTRWVKVISKRPISETRFKVTGLTEGNEYBPH 21729
Qy 138 ----- 137
Db 21730 VMAENAGVGPASGISRLIKCREPVNPPGPTVVKVTDTSKTTVSLEWSKPVFDGMEI 21789
Qy 138 ----- 137
Db 21790 GYIIECKTDLDGMHKVNAEACVKTTRYTVDLQAGEEYKFRVSAINGAGKSDCEVTGTI 21849
Qy 138 ----- 137
Db 21850 KAVDRLTAPELDIDANFKQTHVVVRAGASIRLFIAVQGRPTPTAVMSKPDNSLSRADIHT 21909
Qy 138 ----- 137
Db 21910 TDSFSTLTVCNCRNDAGKYTLTVENNSGSKSITFTVKVLDTPGPPGPTFKDVTGTSAT 21969
Qy 138 ----- 137
Db 21970 LMDAPLDGGARIHHYVVEKREASRRSQVISEKCTROI PKVNDLAEVPPYFRVSAVN 22029
Qy 138 ----- 137
Db 22030 EYGVGEPEMPEPIVATEQAPPRRLDVVDVTSKSSAVLAWLKPDDHSGSRITGYLLEMRO 22089
Qy 138 ----- 137
Db 22090 KGSDLWAEAGHTKQLTFTTVERLVEKTEYEFVRVAKNDAGYSEPREAFSSWIKPEOIEPT 22149
Qy 138 ----- 137
Db 22150 ADLTGNTQLITCKAGSPFTIDVPISGRPAPKVTKLEMLKETDRVSIITTKDRTTLT 22209
Qy 138 ----- 137
Db 22210 VKDSMRGDSGRYFLTLENTAGVKVTSVTVVIGRPGVGTGPIEVSSVSAESCVLWGEPK 22269
Qy 138 ----- 137
Db 22270 DGGTEITNYIVEKRESGTTAMQLVNSSVKRTOIKVTHLTKYMEYFRVSSSENRFVSKP 22329
Qy 138 ----- 137
Db 22330 LESAPIIAEHFPVPPSPAPTPEVYHVHSANAMSIRWEEPHDGGSKIIGYHVEKKERTIL 22389
Qy 138 ----- 137
Db 22390 WVKENKVPCLCNKVTGLVEGLEYPRTYALNAGVSKASEASRRPTMAQNPVDAPGRPE 22449
Qy 138 ----- 137

Db	22450	VTDVTRSTVSLWSAPAYDGGSKVVGYYIERKPVSEVGDGRWLKCNYYTIVSDNFFTTVAL	22509
Qy	138	-----	137
Db	22510	SEGDTYFRVLAKNAAGVISKSESTGPVTCRDEYAPPAELARLHGLDVTIRAGSDLV	22569
Qy	138	-----	137
Db	22570	LDAVGGKPEPKI IWTGDKELDLCEKVSQYTGKATAVIKFCDRSDSGKTYLTVKNAS	22629
Qy	138	-----	137
Db	22630	GTKAVSMVMKVLDSGPCGKLTVSRTQEKCTLAWSLPQEDGAEITHYIVERRETSRLN	22689
Qy	138	-----	137
Db	22690	WVIVEGECTLSYVTVRLIKNNEYIFRVRANKYGPVPESEPIVARNSTFIPSPGP	22749
Qy	138	-----	137
Db	22750	EEVGTGKEHIIQWTKPESDGGNEISNLYVDKREKESLRWTRVNDYVYDTRLKVTSLM	22809
Qy	138	-----	137
Db	22810	EGCDYQFRVTA VNAAGNSEPSESNFISCREPSYTPGPSAPRVVDTTKHSISLAWTKPM	22869
Qy	138	-----	137
Db	22870	YDGGTDIVGVVLEMOEKDQDOWYRVHTNATIRNTEFTVPLKMGOKYSFRVAANVWKMS	22929
Qy	138	-----	137
Db	22930	EYSESAIEPVERIEIPLELADDLKKTVTIRAGASLRMLVSVGRPPPVITWSKQGD	22989
Qy	138	-----	137
Db	22990	LASRAIIDTESYLLIVDKVNRDAGKVTIEAENOSGKSATVLVKVYDTPGPCPSVKV	23049
Qy	138	-----	137
Db	23050	KEVSRDSVTITWEIPTDGGAPINNYIVEKREAAAFKTVTTKCSKTYLRISGLVEGTM	23109
Qy	138	-----	137
Db	23110	HYPRVLPENIYGICEPCESTDVLSVEVPLPAKLEVDVTKSTVLAWEKPLYDGGSL	23169
Qy	138	-----	137
Db	23170	TGYVLEACKAGTERMKVTLKPTVLEHTVTSNEDGEQYLFIRIAQNEKGVSEPRETVTA	23229
Qy	138	-----	137
Db	23230	VTQDLRVLP TIDLSTMPQKTIHPAGRPVELVIPIAGRPPPAASWFFAGSKLRESERV	23289
Qy	138	-----	141
Db	23290	VETHVKAKLTIRETTIRDTGEYLELKNVTGTTSETIKVILDKGPPTGPIKIDEIDA	23349
Qy	142	-----	141
Db	23350	TSITISWEPPELDGGAPLSGVVVEQDARHPGWLPSVESVTRSTFKFTRLTEGNEYVFRV	23409
Qy	142	-----	141
Db	23410	AATNRFIGSYLOSEVIECRSSIRIPGPPELQIFDVSRDGMILTWTYPPEDDGGSQVTGY	23469
Qy	142	-----	141
Db	23470	IVERKEVRADRWVRVNVPTMTTRYRSTGLTEGLEYEHRVTA INARGSGKPSRPSKPIVA	23529
Qy	142	-----	141
Db	23530	MDPIAPGKQPONPRVTDTRTSVLSAWSVPEDEGGSKVTYGILIEQMKVQDHEWTKCNTTP	23589
Qy	142	-----	141
Qy	142	-----	141
Db	23590	TKIREYTLTHLPOGAEYRFRVLACNAGGCEPAEVPGTVKVTEMLBYDPVELDERYOEGI	23649
Qy	142	-----	141
Db	23650	FVRQGGVIRLTIPKIGKPPFICKTWKEGQDISKRAMIATSETHTELVIKADRGDSGYD	23709
Qy	142	-----	141
Db	23710	LVLENCKGKXAVYKVRVIGSPNSPEGPLEYDDIQVRVSRVSMRPADGGADILGYILE	23769
Qy	142	-----	141
Db	23770	RREVPKAAWVTIDSRVRGTSLVVVKGLKENVEYHFRVSAENOFGISPLKSEEPVTPKTPL	23829
Qy	142	-----	141
Db	23830	NPPEPPSNPEVLDVTKSVLSWSRPPKDDGGRVTGYTIERKETSTDKVVRHNKTOITT	23889
Qy	142	-----	141
Db	23890	TMYTGTGLVPDAEYQFRIIAQNDVGLSETSPASEPVVCKDPFKPSQSGELEILSISKDS	23949
Qy	142	-----	141
Db	23950	VTLOWEKEPCDGGKEILGYWVEYRQSGDSAWKKSINKERI KDKOFTIGGLLEATEYEPFRV	24009
Qy	142	-----	141
Db	24010	AENETGLSRPRRTAMSITKLTSGEAPGIRKEMKDVTTLKGEAAQLSCQIVGRPLDKW	24069
Qy	142	-----	141
Db	24070	YRFKELIQSRKYKSSDGRTHLTVMTEDEDEGVYTCIATNEVGEVETSKLLLOATP	24129
Qy	142	-----	141
Db	24130	QHPGYPLKXYGAVGSTLRHLVHYI GRVPAMTWPHGQKLONSENITIENTEHTYHL	24189
Qy	142	-----	141
Db	24190	VKNVORKTHAGYKVKVLSNVFGTVDAILDVEIQDKPKPTGPIVIEALLKNSAVISWKP	24249
Qy	142	-----	141
Db	24250	PADGGSMITNYVVEKCEAEGAEWQLVSSAISVTTCRIVNLTENAGYVFRVSAQNTFGI	24309
Qy	142	-----	141
Db	24310	SDPLEVSSVWIIKSPFEKPGAPGKPTITAVTKQSCVVAWKKPPASDGGAKIRNYLLEKREK	24369
Qy	142	-----	141
Db	24370	KONKWI SVTTEIRETVFSVKNLIEGLEYPFRVKENLGSESEISEPIPTKSDVPIQ	24429
Qy	142	-----	141
Db	24430	APHFKELRNLRVQSNATLVCKVTHPKPIVKWYRQGEI IADGLKYRIOEFKGGYHQ	24489
Qy	142	-----	141
Db	24490	LIASVTDODATVYQVRATNOGGSVSGTASLEVEVPAKIHLPKTLEGMGAVHALRGEVVS	24549
Qy	142	-----	141
Db	24550	IKIPSGKPDPIVTHQKQDLDINNNGHYQVIVTRSFSTSLVPBNVGRKADGAFVVCNAKR	24609
Qy	142	-----	141
Db	24610	FGIDQKTVELDVADVPDPPRGVKVSDASRDSVNLTWTEPASDGGSKITNYIVEKCATTAE	24669

Db 730 HIKTTDQKGMHISSQIKKTTDLTTERLVHVDKRRPTASPHFTVSKI SVPKTEHCYEASIA 789
Qy 19 ----- 18
Db 790 GSAIATLQKELSAISSAQKITKSVKAPTVPKSETRVRAEPTPLPQFPFADPTDYKSEAG 849
Qy 19 ----- 18
Db 850 VEVKKEGVSGITGTVREERFEVLHGREAKVTETARVPVEI PVTPTLVSGLKNTVI 909
Qy 19 -----FQ-----AHQEDTERYV- 30
Db 910 EGESVTLCHISGYPSTVTHYREDYQIESSIDFQITFQSGIARLMIREAFADSGRFTC 969
Qy 31 ----- 30
Db 970 SAVNEAGTVSTCYLAVQVSEFEKETTA VTEKFTTEERKREVRDVMVMTDTSLEEQAQ 1029
Qy 31 ----- 30
Db 1030 PGEPAAPYFITKPVVQKLVGEGSVVFGQVGNPKPHVYWKSGVPLTTGYRYKVSYNKQ 1089
Qy 31 ----- 30
Db 1090 TGECKLVISMTFADDAGEYTIIVRNKHGETSASLSLEADYELLMKSOQEWLYQTQVTA 1149
Qy 31 ----- 30
Db 1150 FVQEPKVGETAPGFVYSEYEKEQALIRKMAKDTVVVRYVEDQEFHISSEERLI 1209
Qy 31 ----- 30
Db 1210 KEIEYRIINTLEELLEDGCEKMAVDISEAVESGFDLRINKYRIILEGMGVTTHCKMS 1269
Qy 31 ----- 30
Db 1270 GYPLPKIANYKDGRIKHGERYQMDFLQDGRASLRIPVVLPEDEGIYTAFA SNIKGNAC 1329
Qy 31 ----- 30
Db 1330 SGKLYVEPAAPLGAPTYIPTLEPVSRISLSPRSVSRSPRMSPARMSPARMSPA 1389
Qy 31 ----- 30
Db 1390 RMSPGRRLEETDESQERLYKPVFLKPVSKFCLEQOTARFDLKVVGRPMPETFWHDQO 1449
Qy 31 ----- 30
Db 1450 QIVNDYTHKVVIKEDGTQSLIIVPATPSDGEWTVVAQNAGRSSISVILTVEAVEHQVK 1509
Qy 31 -----LTNLI----- 36
Db 1510 PMFEVLKKNVNIKEGSOLEMKVRATGNPNPDI VMLKNSDIIVPHKYPKIRIEGTKEAAL 1569
Qy 37 ----- 36
Db 1570 KIDTVSODSAWYATATAINKAGRODTRCKVNVVEFAEPEPERKLIIPRGTYRAKEIAAP 1629
Qy 37 ----- 36
Db 1630 ELEPLHLRYQEQWEGDLYDKEKQKPPFKKXLSLRLKRFQPAHFCECLTPIGDPTWV 1689
Qy 37 ----- 36
Db 1690 VENLHDGKPLEAANRLRMINEFGYCSLDYGVAYS RSDSGIITCRATNKYGTDHTSATLIVK 1749
Qy 37 -----GA----- 38
Db 1750 DEKSLVEESQLEGRKGLQRIEELERMAHEGALTGVTTDQKEKQKPDIVLYPEPVRVLEG 1809
Qy 39 ----- 38
Db 1810 ETARFRCRVTGYPQPKVNMWYNGOLIRKSRFRVRVYDGIHYLDIVDCKSYDTGEVKVTAE 1869

Qy 39 ----- 38
Db 1870 NPEGVTEHKVKLEIQOEDFRSVLRRAPRPEFHVHBPGLQFEVQKVDRPVDTTETKE 1929
Qy 39 -----ELLR--- 42
Db 1930 VVKLKRARI THEKVPESEELRSKPKRRTEGYEAITAVELKSRKKDSEYELLRKTK 1989
Qy 43 ----- 42
Db 1990 DELLHWTKELTBEKKALABEGKITIPTPKDKIELSPSMEAPKIFERIOSQTVOGSDA 2049
Qy 43 -----DPS----- 45
Db 2050 HPRVRVVGKPDPECEWYKNGVKIERSDRIYWYWPEDNVCELVIDVTAEDSASIMVKAIN 2109
Qy 46 ----- 45
Db 2110 IAGETSSHAPLLVQAKOLITFTOELQDVVAKEDTMTTPECETSEPFVKVYKWKDGMVH 2169
Qy 46 ----- 45
Db 2170 EGDYRHMHSRDKVHFLSILITDTSDAEDYSCVLVEDENVKTTAKLIVEGAVVEFVKELQD 2229
Qy 46 ----- 45
Db 2230 IEVPESYSGELECIVSPENIEGKWHYHNDVELKNGKYTITRRGRQNLTVKDVTKEDQGE 2289
Qy 46 ----- 45
Db 2290 YSFVIDGKTTCKLKMKPRPIAILQGLSDQKCEGDIVQEVKVSLESVEGVMMKGOEV 2349
Qy 46 ----- 45
Db 2350 QPSDRVHIVIDKQSHMLLIEDMTKEDAGNYSFTIPALGLSTSGRVSVYSDVITPLKDVN 2409
Qy 46 ----- 45
Db 2410 VIEGTKAVLECKVSPDVTSVKMYLNDEQIKPDDRVOAIVKGTQKORLVINRTHASDEGYP 2469
Qy 46 ----- 45
Db 2470 KLIQVRETNCNL SVEKIKIIRGLRDLTCTETQNVVFEVLSHSGIDVLMNFKDKIKPS 2529
Qy 46 ----- 45
Db 2530 SKYKIEAHGKIYKLTVLNMMKDDGKVTYFAGENITSGKLTVAGGAISKPLTDQTVAESQ 2589
Qy 46 ----- 45
Db 2590 EAVFECEVANPDSKGWLRDQKHLPLTNNIRSESDGHKRRLIITAATKLDDIGEYTYKVAT 2649
Qy 46 ----- 45
Db 2650 SKTSAKLKVEAVKIKKTLKNLTVTOTDAVFTVELTHPNVKGQVQWIKNGVVLESNEKYAI 2709
Qy 46 -----LGAQFRVHL---VQWV----- 58
Db 2710 SVKGTIYSLRIKNCIAVDES VYGFRLGLRGASARLHVETVKI I KKPDKVTALENATVAFE 2769
Qy 59 ----- 58
Db 2770 VSVSHDTPVVKWFHKNVEIKPSDKHRLVSEKRVKJMLQNLSPSDAGEYTA VVGOLECKA 2829
Qy 59 ----- 58
Db 2830 KLFVETLHITKMKNIIEVPETKTASFECEVSHFNVPMSWMLKNGVEIEMSEKPKI VVQCKL 2889
Qy 59 ----- 58
Db 2890 HQLIIMTSTEDSAEYTFVCGNDQVSATLTVTPI MITSMLKDINAEEKDITTFEVTVNYE 2949

```
QY 59 ----- 58
Db 2950 GISYKWLKNGVEIKSTDKCOMRTKLLTHSLINRVHFGDAADYTFVAGKATSTATLYVEA 3009
QY 59 ----- 58
Db 3010 RHIEFRKHIDIKVLEKKRAMFEVSEPDITVQMKDDQELQITDRIKIOKEYVHRL 3069
QY 59 ----- 58
Db 3070 IPSTRMSDAGKYTVVAGGNVSTAKLFVEGRDVRIRSIKKEVOVIEKORAVVEFVNEDDV 3129
QY 59 ----- 58
Db 3130 DAHWYKDGIEINFQOERHKYVVERIRHRMFISETROSDAGEYTFVAGNRSSVTLYVNA 3189
QY 59 ----- 58
Db 3190 PEPPQVLQELQVTVOSGKPARFCAVISGRPOPKISWYKEEQLLSTGFKCKFLHDCQEYT 3249
QY 59 ----- 58
Db 3250 LLLIEAPPDAANYTCEAKNDYGVAATTSASLSVEVPEVSPDQEMPVYPPIITPLQDTV 3309
QY 59 ----- 58
Db 3310 TSEGOPARFCRVSGTDLKVSWYKDKKIKPSRFRFRMTQFEDTYQLEIAEAYPEDEGTYT 3369
QY 59 ----- 58
Db 3370 FVASNAVQVSSANLSLEAPESILHERIEOEIEMEMKEFSSSFLSABEEGLHSAELQLS 3429
QY 59 ----- 58
Db 3430 KINETLELLSESPVYSTKFDSEKEGTGPIFIKEVSNADISMGDVATLSVTVIGIPKPIQ 3489
QY 59 ----- 58
Db 3490 WFFNGVLLTPSADYKVFVGDGDDHSLIILFTKLEDEGEYTCNASNDYGKTCISAYLKINSK 3549
QY 59 ----- 58
Db 3550 GEGHKOTETESAVAKSLEKLGCPGPPHFLKELKPIRCAOGLPAIFEYTVVGEPAPTVWF 3609
QY 59 ----- 58
Db 3610 KENQOLCTSVYVYTIHNPNGSGTFIVNDPQREDSGLYICKAENMLGSESTCAAEILLVLED 3669
QY 59 ----- 58
Db 3670 TDMTDTCKAKSTPEAPEDFPQTPKGPVAVBALDSEQEIATFVKDTILKAALITEENQOL 3729
QY 59 ----- 58
Db 3730 SYEHIKANELSSQLPLGAQELQSILODKLTPESTREFLCINGSIHFPQKPESPNQL 3789
QY 59 ----- ILTEPEG- 65
:
:
:
:
Db 3790 QIVSQKTSKXEGILMPPEPETQAVLSDTEKIFPSAMSIEQINSITVEPLKTLAEPEGN 3849
QY 66 ----- 65
Db 3850 YPOSSIEPPMHSYLTSVAEVLSPKEKTVSDTNREQRTVLOKQEAQSALILSOLAEHV 3909
QY 66 ----- 65
Db 3910 ESLOSFDVMIQVNYEPLVPSHSECTEGGKILIESANPLENAGQDSAVRIEKGSLRPPL 3969
QY 66 ----- 65
Db 3970 ALEEKVLLKEHSDNVMPDQIIIESKREPVAIKKVOEQVRDILLSKESLLSGIPESQR 4029
QY 66 ----- 65

Db 4030 LNLKIQICRALQAAVASEQPGOLFSEWLRNIEKVEAVNITQEPRHIMCMYLVTSKSVT 4089
QY 66 ----- 65
Db 4090 EEVTHIIEDVDPMANLQWELDALCAIYBEEIDILTAEGPRIQOGAKTSLQEEMDSFG 4149
QY 66 ----- 65
Db 4150 SQKVEPITEPEVESKYLISPEEVSFVQSRVKYLDATPTKGVASAVVSDKQDES LKP 4209
QY 66 ----- 65
Db 4210 SEEKESSSESGTEBVA TVKIQEAGGFIKEDGPMIHTPLVDTVSEEGDI VHLTTSITNA 4269
QY 66 ----- 65
Db 4270 KEVNMWFENKLVPSDEKFKCLQDQNTYTLVIDKVNTHDHOGEYVCEALNDSGKTATS AKL 4329
QY 66 ----- 69
:
:
:
:
Db 4330 TVVKRAAPVIRKRIEPLLEVALCHLAKFTCEIOSAPNVRFQWFKAGREIYESDKCSIRSK 4389
QY 70 ----- 69
Db 4390 YISSLEILRTQVDCGEYTCASNEYSVSVCTATLTVTVPGEKKVKLLPERKPEPKEE 4449
QY 70 ----- 69
Db 4450 VVLKSVLRRRPEEBEPKVEPKKLEKVKPAVPEPPPKPVEEVPVTVTKRERKIPEPTK 4509
QY 70 ----- 69
Db 4510 VPEIKPAIPLPAPEPKPKPEAEVKTIKPPVPEPEPTIAAPVTVPVGKAEAKAPKEA 4569
QY 70 ----- 69
Db 4570 AKPKGPIKGVPKTSPTEAERRKLRPGSGGKPPDEAPFTYQLKAVPLKFVKEIKDIL 4629
QY 70 ----- 69
Db 4630 TESEFVGSSAIFELVSPSTAITTWMKDSNIRESPKHFIAADGDKRKHIIIDVQLSDAG 4689
QY 70 ----- 69
Db 4690 EYTCVLRGNKEKSTAKLVVEELPVRVFKTLEEEVTVVKGQPLYLSCELNKERDVVRWK 4749
QY 70 ----- 69
Db 4750 DGKIVVEKGRIVPGVIGLMRALTINDADDTDAGTVTVTVENANNLECSVVKVVEVIRD 4809
QY 70 ----- 69
Db 4810 WLVKPIRDQHVKPGKTAIFACDIADKTPNIKWFPGYDEIPAEPPNDKTEILRDGNHLYLKI 4869
QY 70 ----- 69
Db 4870 KNAMPEDIAEYAVEIESGRYPAKLTUGEREVELLKPIEDVTIYKESASFDAEISADIP 4929
QY 70 ----- 69
Db 4930 GOWKLXGELLRPSPTCEIKAEGGKRFLLRKVKLDOAGEVLYOALNAITTAILTIVKEIEL 4989
QY 70 ----- 69
Db 4990 DFAVPLKDVTVPERROARFECVLTREANVWSKGPDI IKSSDKFDI IADGKHHILVINDS 5049
QY 70 ----- 69
Db 5050 QFDDGEVYTAEVGKKT SARL FVTGIRLKFMSPLEDQTVKEGETATFVCELSHEKMHVW 5109
QY 70 ----- 69
```

Db	5110	FKNDAKLHTRVTLISSEKTHKLEMKVTLDDISQIKAQVKELSTAQLKVLADPYFT	5169
Qy	70	-----	69
Db	5170	VKLHDKTAVEKDEITLKCVSKDVPVKWFKDGEBEIVPSPKYSIKADGLRRILIKKADIK	5229
Qy	70	-----	69
Db	5230	DKGEYVDCGDTKANKVTVEARLIKVEKPLYGVFVGETAHFEIELESEPDVHCWKLK	5289
Qy	70	-----	69
Db	5290	GQPLTASPDCEIIEGCKHLILILHNCQLGWTGEVSFOAANAKSAANLKVKELPLIFITPL	5349
Qy	70	-----	69
Db	5350	SDVKVEKDEAKPECEVSREPKTFRWLKGTQETGDDRFELIKDGTKHSMTKSAAFEDE	5409
Qy	70	-----	69
Db	5410	AKYMFADKHTSGKLIIEGIRLKFUTPLKDVTAKEKESAVFTVELSHDNIRVKWFKNDQ	5469
Qy	70	-----	69
Db	5470	RLHTRSVSMQDEGKTHSITFKDLSIDDTSQIRVEAMGHSSEAKLTVEGDPYFTGKLQD	5529
Qy	70	-----	69
Db	5530	YTGVKEDEVILCEISKADAPVKWFKDGEIKPSKNAVIKADGKRMLILKALKSDIQ	5589
Qy	70	-----	69
Db	5590	YTCDCGTDKTSGLDIEDREIKLVRPLHSVEVMEETARFETISEDDIHANWKLGEAL	5649
Qy	70	-----	69
Db	5650	LQTPDCEIKEGKIHLVHLNCRLDQGGVDFQAANVKSASHLRVKRPRVIGLRPLKDV	5709
Qy	70	-----	69
Db	5710	VTAGETATFDCELSYEDIPVEWYLGKKEPSPKVPVRSSEKGVHTLTLRDVKLEDA	5769
Qy	70	-----	69
Db	5770	LTAKDFKTHANLKVKEPPVEFTKPLEDOTVEEGATAVLECEVSRENKVKWFKNGTEILK	5829
Qy	70	-----	69
Db	5830	SKKYEIVADGRVKRLVIHDCTPEDIKTYTCDAKDFKTCNLNVPPHVEFLRPLDLQVR	5889
Qy	70	-----	69
Db	5890	EKENARFECELSRENKVKWFKDGAETKGGKYDIISKGAVRILVINKCLLDDEAEYSCE	5949
Qy	70	-----	69
Db	5950	VRTARTSGLTVLEEAFTKLANIEVSETDTIKLVCEVSKPGAEVIWYKGDDEIIEG	6009
Qy	70	-----	69
Db	6010	RYEILTEGRKILVIONAHLEDAGNYNCRPLSPSRDTGKVKVHELAAEFISKPNLELEG	6069
Qy	70	-----	69
Db	6070	EKAEFVCSISKESFPQWKRDDKTLES GDKYDVIADGKKRVLVVKDATLQDMGYVMVG	6129
Qy	70	-----	69
Db	6130	AARAAHLTVIEKLRIVVPLKDRVKEQEVVFNCEVNTGAKAKWFRNEAIFDSSKYI	6189
Qy	70	-----	69
Db	6190	ILOKDLVYTLIRDAHLDDQANYVSLNHRGENVKSAANLIVEEEDLRIVEPLKDIETM	6249
Qy	70	-----	69
Db	6250	EKKSVTWCKVKNRLNVLTKWTQNGEEVFPDNHRSYRVVDKYKHMLTIKDCGPPDSEGI	6309
Qy	70	-----	69
Db	6310	AGQDKSVAELLIIIEAPTEFVEHLEDQTVTFPDDAVFSCOLSREKANVKWYRNGREIK	6369
Qy	70	-----	69
Db	6370	KYFEKOGSIHRLIIKOCRLDDECEYACGVEDRKSRARLFEVEIPVEIIRPPQDILEAP	6429
Qy	70	-----	69
Db	6430	ADVPLAELNKDKVEQWLRNNMVVQGDQKHQMSEKGIHRLQICDIKPRDQGEYRFAK	6489
Qy	70	-----	69
Db	6490	DKEARAKLEAAPKIKITADQDLVVDVGKPLTMVVPYDAYPKAEAEWPKENEPLSTKTID	6549
Qy	70	-----	69
Db	6550	TTABQTSFRILEAKKGDKGRYKIVLQNHGKAGFINLKVDPGVRNLEVTFETFDGEV	6609
Qy	70	-----	69
Db	6610	SLAWEEPLTDGSKIIGYVVERRDIKRKTWVLTADRAESCEFTVTGLOKGVVFLRVSA	6669
Qy	70	-----	69
Db	6670	RNRVGTGPEVETDNPVEARSKYDVPGLNVITITDVRNRFVSLTWEPPEYDGGAEITNV	6729
Qy	70	-----	69
Db	6730	IELRDKTSIRWDTAMTVRAEDLSATVTDVVEGQEYSFRVRAQNRIGVCKPSAATPFVKA	6789
Qy	70	-----	76
Db	6790	DIERPSPVNLTSDDQTSQSSVOLKWEPLKDGSGPILGYIIERCEGKDNMIRCNMKLV	6849
Qy	77	-----	76
Db	6850	PELTYKVTGLEKGNKYLYRVAENKAGVSDPSEILGPLTADDAFVEPTMDLSAFKDGLEV	6909
Qy	77	-----	76
Db	6910	IVPNPITILVPSTGYPRPTATWCFQDKVLETCGRVVKMKTLSAYAEVLISPSERSDKGI	6969
Qy	77	-----	76
Db	6970	LKLENRVKTSIGEIDVNVVIARPSAPKELKFGDITKDSVHLTWEPDDDDGSGPLTGYVVEK	7029
Qy	77	-----	81
Db	7030	REVSRTWTKVDMFVTDLEFTVPDLVOGKEYLFKVCARNKCGGPEPAYVDEPNVMTPT	7089
Qy	82	-----	81
Db	7090	VDPPEVNVKWRDRTANSIFLTWDPKNDGSRIGYIVERCPRGSKWVACGEPVAETKM	7149
Qy	82	-----	81
Db	7150	EVTGLEGKWAYRVKALNRQAGSKPSRPTBEOAVDTQEAPEIFLDVKLLAGLTVKACT	7209
Qy	82	-----	81
Db	7210	KIELPATVTKPEPKITWTKADMILKQDKRITENVPKSTVTIIVDSKRSDTGTIIEAV	7269
Qy	82	-----	81
Db	7270	NVCGRATAVVEYNVLDKPGPAAFDITDVTNESCULLTWNPPRDGSGKITVYVVERRATD	7329

QY	82	-----	81
Db	7330	SEVHKLSSVTNFKATKLIENKEVIFRVAENMYGCEPVQASPIITAKYQFDPGPP	7389
QY	82	-----	81
Db	7390	TRLEPSDITKDAVLTWCEPDDGGSPITGYWVERLDPTDKWRCNKMPVKDITTVRVKG	7449
QY	82	-----	81
Db	7450	LTNNKKYFRVLAENLAGPKSPKSTEPILIKOPIDPPWPKPTVKDVGKTSVRLNWK	7509
QY	82	-----	81
Db	7510	PEHGGAKIESYVIEMLKTDGVRVAEGVPTTQHLLPGLMEGQESFRVRAVNKAGES	7569
QY	82	-----	81
Db	7570	EPSESPDLCREKLYPPSPRMLVINITKNTADLKWTVPKDDGGSPITNYIVEKRDVR	7629
QY	82	-----	81
Db	7630	RKGHOTVDTTKCTVTPLTEGSLYFRVAENAIQSDYTEIEDSVLAKDTFTTPGP	7689
QY	82	-----	81
Db	7690	PYALAVVDVTKRHVDLKWEPKNDGGRPIORYVIEKKERLGRWVKAGTAGPDCNFRVT	7749
QY	82	-----	81
Db	7750	DVIEGTEVOFVRAENAGVHPSEPTIELSIDPTSPSPPLDLHVTDAGRKHIAIAWK	7809
QY	82	-----	81
Db	7810	PPEKNGSPILGYHVEMCPVTEKMRVNSRPIKDLKFVEGVVPDKEYVLRVAVNAI	7869
QY	82	-----	81
Db	7870	GVSEPSEISENVAKDPCKPTIDLETHDIIIEGKLSIPVPFRAVPVTVSWHKOGKE	7929
QY	82	-----	81
Db	7930	WKASRLTMKNDHSAHLEVPKSVRADAGIYITILENKLGSATASINVKVIGLPGPCKDI	7989
QY	82	-----	81
Db	7990	KASDITKSSCKLTWEPPEFDGTPILHYVLERREAGRRTYIPVMSGENKLSWTVKDLIPN	8049
QY	82	-----	81
Db	8050	GEYFFRVKAVNKVGGEVIELKNPVIAQDPKQPPDPPVDVEVHNPTAEAMTITWKPPLYD	8109
QY	82	-----	81
Db	8110	GGSKIMGYIIIEKIAGEERWKCNEHLVPILTYTAKGLEEGKEYQFVRAENAGISEPS	8169
QY	82	-----	81
Db	8170	RATPPTKAVDPIDAPKVILRTSLEVKRGDEIALDASISGSPYPTITWIKDENVIIVPEIK	8229
QY	82	-----	81
Db	8230	KRAAPLVRRRKGVEBEPFVLPLTORLSIDNSKKGESQLRVRDSLDPDHGLYMIKVEND	8289
QY	82	-----	81
Db	8290	HGIAPACTVSLDTPGPINFVFEDIRKTSVLCKWEPLDDGGSEIINYTLEKKDKTKP	8349
QY	82	-----	81
Db	8350	DSEWIVVTSTLRHKYSVTKLIEGKEYLFRVRAENRFGPPPCVSKPLVAKDPFGPPDAP	8409
QY	82	-----	81

Db	8410	DKPIVEDVTSNMLVKWNEPKDNGSPILGYWLEKREYNSTHWSRVNKSLLNALKANVDGL	8469
QY	82	-----	81
Db	8470	LEGLTYVFRVCAENAAAGPKFSPSPDKTAHDPISPQPPPIPRVTDTSSTTIELEWEPPA	8529
QY	82	-----	81
Db	8530	FNGGGEIVGVFVDKOLVGTNEWSRCTEKMIKVRQYTVKEIREGADYKLRVSAVNAAGSGP	8589
QY	82	-----	81
Db	8590	PGETOPVTVAEPOEPPAVELDVSVKGGIQIMAGKTLRIPAVVTVGRPVPTKVMTKEGELD	8649
QY	82	-----	81
Db	8650	KDRVVIDNVGTSKELI IKDALRKDHGRYVITATNSCGSKFAAARVEFVDFVPGVLDLRPV	8709
QY	82	-----	81
Db	8710	VTRKXKLLNWDSDPEDGGSEITGFIIERKDAKMTWQPIETERSKCDITGLLEGQEVK	8769
QY	82	-----	81
Db	8770	FRVIAKNKFGCGPPVEIGPILAVDPLGPPTSPLRITYTERTKSTITLDWKEPRSGGSP	8829
QY	82	-----	81
Db	8830	QGYIIIEKRHDKPDFERVNKRCLPTTSFLVENLDEHOMYEFVRKAVNEIGESEPSPLNV	8889
QY	82	-----	81
Db	8890	VIQDDEVPTIKLRLSVRGDTIKVKAGEPVHIPADVTLGLMPKIEWSKNETVIEKPTDAL	8949
QY	82	-----	81
Db	8950	QITKEVSRSEAKTELSIPKAVREDKGTVTASNRLSGVFRNVHVEVDPRSPRNLA	9009
QY	82	-----	81
Db	9010	TDIKAESCYLTWDAPLDNGGSEITHYVIDKRDAKRKAEEVETAVKRYGIWKLIPN	9069
QY	82	-----	81
Db	9070	GQYEFVRVANKYGISDECKSKVVIQDPVRLPGPGPKVLARTKGSMLVSWTPTLDNG	9129
QY	82	-----	81
Db	9130	GSPITGYWLEKREEGSPYMSRVSRAPITTKVGLKGVFNVRLLLEGVKYQBAMAINAAGI	9189
QY	82	-----	83
Db	9190	GPPSESPDEVAGDPIFPPOPSPCEVKDKTKSSISLGHKPPAKDGGSPKIGYIVEMQEE	9249
QY	84	-----	83
Db	9250	GTTDKRVNEPDKLITTCCECVVNLKELRKYRFRVKAVNEAGESEPSDTTGEIPATDIOE	9309
QY	84	-----	83
Db	9310	EPEVFIGAQDCLVCKVAGSQIRIPAVIKGRPTPKSSWEPFGKAKAMKGVHDIPEDAQ	9369
QY	84	-----	83
Db	9370	LETAENSSVIIIECKRSHTGKYSITAKNAGQKANTANCRVKVMDVPKPKDLKVSIDITRG	9429
QY	84	-----	83
Db	9430	SCRLSWKMPDDGGDRIKGYVIEKRTIDGKAWTKVNDGCGSTTFVVPDLLEQQYFFVR	9489
QY	84	-----	83

Db	9490	AENRFGICPPVETIQTARTARDPIYPPDPPIKIKIGLITKNTVHLSWKPKNDGGSPVTHY	9549
Qy	84	-----	83
Db	9550	IVECLAMDPTGCKKEAMQCNKRDVEELQFTVEDLVGEGYEFVRKAVNAAGVSKPSATV	9609
Qy	84	-----	83
Db	9610	GPCDCQRPDMPPSIDLKEFMEVEBGTNNIVAKIKGVPPFTLTWFKAPPKPDNKEPVLV	9669
Qy	84	-----	83
Db	9670	DTVHKNLVDDTCTLVIFQSRSDTGLYTITAVNNLTASKEMRLNVLGRPPVGPPIKF	9729
Qy	84	-----	83
Db	9730	ESVSADOMTSLWFFPKODGSKITNYVIEKREANRKTWVHVSSEPKECTYTIIPKLEGE	9789
Qy	84	-----	88
Db	9790	YVFRIMAQNKYIGIPELDSEPETARNLFSVPCAPDKPTVSSVTRNSMTVNWEEPEYDGS	9849
Qy	89	-----	88
Db	9850	PVTGYWLEMKDTSKRKRVNRDPIKAMTLGVSVKVTGLIEGSDYQFRVYAINAAGVGA	9909
Qy	89	-----	88
Db	9910	SLPSPDATARDPIAPPPPPKVTDMTKSSADLEWSPLKDGSKVTGYIIVEYKEEGKEE	9969
Qy	89	-----	88
Db	9970	WEKGKKEVRGKLVVTGLKEGAFYKFRVSAVNIAGIGEPCEVTDVIEKDRLVSPDLQL	10029
Qy	89	-----	88
Db	10030	DASVRDIWVHAGVIRIIAYVSGKPPPTVWNMNERTLPOBATTIETTAISSMVINKCQ	10089
Qy	89	-----	92
Db	10090	RSHOGVYSLAKNEAGERKKTIIVDVDPGVGTPFLAHLNLTNESCKLTFSPEDDGS	10149
Qy	93	-----	92
Db	10150	PITNVYIEKESDRRAWTPVTYVTRQNATVQGLIQKAYFERIAAENSIGMGPVETSE	10209
Qy	93	-----	92
Db	10210	ALVIREPITVPERPEDLEVKEVTQNTVTLTNPPKYDGGSEIINYVLESRLIGTEKFHKV	10269
Qy	93	-----	92
Db	10270	TNDNLSRKYTVKGLKEGDTYEVRSVAVNIVGGKPSFCTKPICTCKDELAPPTLHDFRD	10329
Qy	93	-----	92
Db	10330	KLTVRGEAFALTGRYSGKPKVSWFKDEADVLEDDRTHIKTTPATLALALEKIKARSDS	10389
Qy	93	-----	92
Db	10390	GKYCVVVENSTGSRKGFQVNVVDRPVPVGPVSFDEVTQDYVVISWKPPDDGGSKITN	10449
Qy	93	-----	92
Db	10450	YIIEKKEVGKDVMPVTSASAKTTCKVSKLLEKDYIFRIHAENLYGISDPLVSDSMKAK	10509
Qy	93	-----	92
Db	10510	DRFRVPADQPPIVTEVTKDSALVTWNKPHDGGKPTNYILEKRETMSKRWARVTKDPIH	10569
Qy	93	-----	92
Db	10570	PYTKFRVPDLLEGQCFRVSARENIGDIPSPSPKPVFAKDFIAPKPSPPVNPDEADTTTC	10629

Qy	93	-----	92
Db	10630	NSVDLTWQPRHDGSKILGYIVEYQKVGDEEWRANHTPESCPETKYKVTGLRQOTYK	10689
Qy	93	-----	97
Db	10690	FRVLAVNAAGESDPAHVPEPVLVKORLEPPELILDANMAREQHIKVGDTLRLSAIIGVP	10749
Qy	98	-----	97
Db	10750	FPKVTWKEDRDAPTAKRIDVTPVGSKLEIRNAAHEDGGIYSLTVENPAGSKTVSVKVLV	10809
Qy	98	-----	97
Db	10810	LDPGPPROLEVSEIRKDCYLTKWKEPLDDGGSVITNYVVERRDVASAQWSPLSATSKKK	10869
Qy	98	-----	97
Db	10870	SHFAKHLNEGNQYLFRVAENQYGRGFFVETPKPIKALDPLHPPGPKDLHHVDVDKTEV	10929
Qy	98	-----	97
Db	10930	SLVWNPDRDGGSPITGYLVEYQEEGTQDWIKFVTNLECVVTGLQOGKTYRFRVKAEN	10989
Qy	98	-----	97
Db	10990	IVGLGLPDTTPIECOEKLVPPSVELDVKLEGLVVKAGTVTFPAIIRGVPPVPTAKWTT	11049
Qy	98	-----	97
Db	11050	DGSEIKTDEHYTVETDNFSSVLTIKNCLRRDTGEYQITVSNNAAGSKTVAVHLTVLDVPGP	11109
Qy	98	-----	97
Db	11110	PTGPNILDVTPEHMTISMQPKDDGSPVINIVEKQDTRKDTGWVSSGSSKTLKIP	11169
Qy	98	-----	97
Db	11170	HLQKGEYVFRVAENKIGVPLDSTPTVAKHKFSPSPGPKPVVTDITENAAATVSWTL	11229
Qy	98	-----	97
Db	11230	PKSDGSGPITGYMYMERREVTGKVRVNKTPADLKFRVTGLYEGNTYEFVRFAENLAGLS	11289
Qy	98	-----	101
Db	11290	KPSSSDPIKACRPKPPGPPINPKLKDRETDALVMTKPLSDGSGPILGYVVECOKPG	11349
Qy	102	-----	101
Db	11350	TAOWNRINKOBLIRQCAFVRPGLIEGNEYFRIKAANIVGEGEPRELAESVIAKDILHPP	11409
Qy	102	-----	101
Db	11410	EVELDVTCDRVTIVRVGOTIRILARVGRPEPDITWTKEGKVLVREKRVLDLQDLPRVEL	11469
Qy	102	-----	101
Db	11470	QIKEAVRADHGKYIISAKNSSGHAQSAIVNVLDPRPGCONLKVNTVTKENCTISWENPL	11529
Qy	102	-----	101
Db	11530	DNGSEITNFIVEYRKNQKQWSIVASDVTKRLIKANLLANNEYFRVCAENKVGVGPTI	11589
Qy	102	-----	101
Db	11590	ETKTPILAINPIDRGPENLHIADKGKTFVYLVKWRRPDYDGGSPNLSYHVERRLKGSD	11649
Qy	102	-----	101
Db	11650	WERVHKGSIKETHVMVDRCVENQIYEFVRQTKNEGGSMDWKTBEVVVKEDLQKPVLDLK	11709

Qy 102 ----- 101
Db 11710 LSGVLTAKGDTIRLEAGYRGKPFPEVAMTKOKDATDLTRSPRVKIDTRADSSKFLTKA 11769
Qy 102 ----- 101
Db 11770 KRSDBGKVVVATNTAGSFVAVATNVNLDKPGFVRNLKIVDVSSDRCTVCWDPPEDDGGC 11829
Qy 102 ----- 101
Db 11830 EIQNYILEKCTKRMWSTYSATVLTGCTTVTRLIEGNEYIFRVRAENKIGTGPTTESKP 11889
Qy 102 ----- 101
Db 11890 VIAKTYDKGRDPPEVTKVSKSEMTVVMNPPEYDGGKSIITGYFILEKKEKHSTRWVPV 11949
Qy 102 ----- 101
Db 11950 KSAIPERRMKVQNLPLDPEHYQFRVKAENEIGIGESLPSRPPVAKDPIEPGPPTNFRVV 12009
Qy 102 ----- 101
Db 12010 DTTKHSITLWGKVPYDGGAPIIGYVVMRPKIADASPDEGWRKCNAAQLVRKEFTVTS 12069
Qy 102 ----- 101
Db 12070 LDENQEYFRVCAQNVGIGRPAELKEAIKPKBILEPPEIDLDASMRKLVIVRAGCPRL 12129
Qy 102 ----- 101
Db 12130 FAIVGRPAPKVTRKVGIDNVVRKQVDLVDWAFVLPINSTRDDSGKSLTLVNPAGE 12189
Qy 102 ----- 101
Db 12190 KAVFVNVRLDTPGVSUVDKVSVDVTKTCHVSWAPPENDGGSQVTHYVEKREADRKTWS 12249
Qy 102 ----- 101
Db 12250 TVTPEVKKTSFHVNLVPGNEYFRVTAVNEYGFGVTDVPKPVLASDPLSEDPKPKLE 12309
Qy 102 ----- 101
Db 12310 VTEMKNSATLAWLPPLRDGAKIDGYITSYREEEQPADRWTEYSVVVKDLSLVVTGLKEG 12369
Qy 102 ----- 101
Db 12370 KKYKFRVAARNVAVGSLPREAEGVYEAKEQLLPKILMPEQITIKAGKCLRIEAHVYGKP 12429
Qy 102 ----- 101
Db 12430 HPTCKWKGEDEVVTSSHVAHKADSSSILIKDVTRKDSGYSLTAENSSGTDQKIKV 12489
Qy 102 ----- 101
Db 12490 VMDAPGPPQPPFDISDIDADACSLSWHI PLEDGGSNITNYIVEKDVSRGDMVLTALASV 12549
Qy 102 ----- 101
Db 12550 TKTSRVRGLIPQGEYIFRVRAENRFGISEPLTSPQWVAQFPFGVPSEPKNARVTKVND 12609
Qy 102 ----- 101
Db 12610 CIFVAMDRPDSGGSPIIIGYLIERKERNLSLLWVKANDTLVRSTEYPCAGLVEGLEYSFRI 12669
Qy 102 ----- 101
Db 12670 YALNAGSSPPSKPTEYVTARMPDVPKGPEVIDVTKSTVSLIWARPKHDGSKIIIGYFV 12729
Qy 102 ----- 101
Db 12730 EACKLPDGKWRNTAPHQIPOEYATATGLEEKAQYQFRAIARTAVNISPPSESPDPVTI 12789
Qy 102 ----- 101

Db 12790 LAENVPPRIDLSVAMKSLLTIVKAGTNVCLDATVFGKPMPTVSWKKDGTLLKPAEGIKWAM 12849
Qy 102 ----- 101
Db 12850 QRNCTLELFSVNRKSDGYTITAENSNGSKSATIKLVLDKPPASVINKMYSDRAM 12909
Qy 102 ----- 101
Db 12910 LSWEPPELGGSEITNYIVDKRETSRPNWAQVSATVPITSCSVEKLEIEGHEVQFRICAEN 12969
Qy 102 ----- 101
Db 12970 KYGVGDPVFTEPAIAKNPYDPPGRCDPVISNITKDHMTVSWKPPADGGSPITGYLLEK 13029
Qy 102 ----- 101
Db 13030 RETQAVNWKVNRXPIIERTLKATGLOEGTEYEFPRVTAINKAGPKGKSDASKAAYARDPO 13089
Qy 102 ----- 101
Db 13090 YPPAPPAPPKVYDTRSSVSLSWGKPAYDGGSPIIGYLVEVKRADSDNWRNCLPNQLQK 13149
Qy 102 ----- 101
Db 13150 TRFEVTGLMEDTQYQFRVYAVNKGISDSDVPDKHYPKDILIPPEGELDADLRKTLILR 13209
Qy 102 ----- 101
Db 13210 AGVTMRLYVPVKGRPPPKITWSKPNVNLDRIGLDIKSTDFTDFLRCENVNKYDAGKYIL 13269
Qy 102 ----- 101
Db 13270 TLENSCGKKEYTIVVKVLDTPGPPVNVTVKEISKDSAYVWTEPPIIDGGSPINIYVQKR 13329
Qy 102 ----- 101
Db 13330 DAERKSWSTVTTECSKTSFRVANLEEGKSYFRVFAENYIGIDPGETRDVAKASQTPGP 13389
Qy 102 ----- 101
Db 13390 VVDLKVRSVSKSSCSIGHKKPHSDGSRIGYVVDFTLEENKQWRVMKSLSLQYSAKDLT 13449
Qy 102 ----- 101
Db 13450 EGKEYTFRVSAENENGEGTPEITVVARDVVAPDLKGLPDLCLYAKENSNFRUKIPI 13509
Qy 102 ----- 101
Db 13510 KGPAPSVSWKXGEDPLATDTRVSVESSAVNTTLIVYDCQKSDAGKYITIKNVAGTKEG 13569
Qy 102 ----- 101
Db 13570 TISIKVVGKPGIPTGPIKFDEVTAEAMTLKAPPKDDGGSEITNYILEKRDSVNNKWVTC 13629
Qy 102 ----- 101
Db 13630 ASAVQKTTFRVTRLHEGMEYTFRVSAENKYGVGEGLKSEPIVARHFPDVPDAPPPNIVD 13689
Qy 102 ----- 101
Db 13690 VRHDSVSLTWDPKKTGGSPITGYHLEFKERNLSLLWKRANKTPIRMKDFKVTGLTEGLEY 13749
Qy 102 ----- 101
Db 13750 EFRMAINLAGVKPSLPSEPVVALDPIPPGKPEVINITRNSVTLIWTEPKYDGGHKL 13809
Qy 102 ----- 101
Db 13810 GYIWEKRDLPKSKWMKANVNVPECAFTVTDLVEGGKYEFIRAKNTAGISAISESTET 13869
Qy 102 ----- 101

Db 13870 IICKDEYEAPTIVLDPTIKDGLTIKAGDTIVLNAISILGKPLPKSSWSKAGKDIRPSDIT 13929
 Qy 102 ----- 101
 Db 13930 QITSTPTSSMLTIKYATRKDAGEYTTIATNPGTKVEHVKVTVLDVPGPPGVEISNVSA 13989
 Qy 102 ----- 101
 Db 13990 EKATLTWTPLEDGSGPIKSYILEKRETSRLLTWTVVSEDIQSCRHVATKLIQONEVIFRV 14049
 Qy 102 ----- 101
 Db 14050 SAVNHVKGEPVQSEPVQKMDRFGPPGPEKPEVSNVTNTATVSWKRPVDDGSGSEITGY 14109
 Qy 102 ----- 101
 Db 14110 HVERREKKSIRWRAIKTPVSDLRCKVTGLQEGSTVEFRVSAENRAGIGPPSEASDVL 14169
 Qy 102 ----- 101
 Db 14170 KDAAYPPGPPSNPHVDTTCKKSASLANGKPHYDGLLEITGYVVEHOKVGDEAWIKDTTGT 14229
 Qy 102 -LYITRF----- 107
 Db 14230 ALAITQFVWPDLOTKEKYNFRISAINDAGVGEPVDPDVEIVEREMAPDFELDAELRTL 14289
 Qy 108 ----- 107
 Db 14290 VVRAGLSIRIFVPIKGRPAPEVTWTKONINLKNRANIENTESFTLLIIPECNRYDTGKFV 14349
 Qy 108 ----- 107
 Db 14350 MTIENPAGKSGFVNVVRVLDTPGPNLNRPTDITKDSVTLHMDPLIDGSGRITNIVEK 14409
 Qy 108 ----- 107
 Db 14410 REATRKSYSTATTCKCHKTYKVTGLSGCEYFFRVMAENEYEGICEPTETTEPVKASEAPS 14469
 Qy 108 ----- 107
 Db 14470 PPSLNIMDITKSTVSLAWPKPKHGGSKITGYVIEAQRKSGDQWTHITTVKGLECVVRN 14529
 Qy 108 ----- 107
 Db 14530 LTEGEETTFQWMAVNSAGSAPRESRPVIVKEQTMPELDLRGIYQKLVIAKAGDNKVE 14589
 Qy 108 ----- 107
 Db 14590 IPVLRPKPTVTKKGQILKQTORVNFETTATSTILNINECVRSDSGPYPLTARNIVGE 14649
 Qy 108 ----- 107
 Db 14650 VGDVITIQHDIIPGPTGPIKFEVSDVFTFSWDPDENGGVPISNVVEMRQDSTTW 14709
 Qy 108 ----- 107
 Db 14710 VELATTVIRTTYKATRLTTGLEQYQVRVKAQNRGVGPGITSACIVANYPFKVPGPGTPQ 14769
 Qy 108 ----- 107
 Db 14770 VTAVTKDSMTISWHEPLSDGGSPILGYHVERKERNGILMOTVSKALVPGNIFKSSGLTDG 14829
 Qy 108 ----- 107
 Db 14830 IAYEFRVIAENMAGSKSPSEPMALDIPDPPOKVPPLNITRHTVTLKWAKPEYTGGF 14889
 Qy 108 ----- 107
 Db 14890 KITSYIVEKRDLPNGRWMLKANFSNILENEFTVSGLTDAAYEFRVIAKNAAGAISSPPSEP 14949
 Qy 108 ----- 107
 Db 14950 SDAITCRDDVEAPKIKVDVKFKDTVILKAGEAFRLADVSGRPPPTMWSKDGKELECTA 15009

Qy 108 ----- 107
 Db 15010 KLEIKIADFSTNLVNDKSTRDRSGAYTLTATNPGGFAKIFNVKVLORPGPPEGLAVTE 15069
 Qy 108 ----- 107
 Db 15070 VTSEKCVLSWFPPLDDGAKIDHYIVQKRETSRLAWTNVASEVQVTKLVTKLLKGNEYI 15129
 Qy 108 ----- 107
 Db 15130 FRVMAVNYGVGEPLSEPLESEPLAVNPYPGPPPKNPEVTTITKDSMVVCGHPDSDGSEI 15189
 Qy 108 ----- 107
 Db 15190 INYIVERRDKAGORWIKCNKKTLDRLYKVSGLTEGHEYEFRIEMAENAGISAPSTSPF 15249
 Qy 108 ----- 107
 Db 15250 YKACDTVPKPPGPNRVLDTSRSSISIAMNKPIYDGGSEITGYMVEIALPEDEMQIVT 15309
 Qy 108 ----- 107
 Db 15310 PPAGLKATSYITITGLTENQEKIRIYAMNSEGLGEPALVPGTPKAEDRMLPPEIBDADL 15369
 Qy 108 ----- 107
 Db 15370 RKVVTIRACCTLRFPVPIKGRPAPEVKWARDHGESLDSKASIESTSSYTLIIVGNVNRFD 15429
 Qy 108 ----- 107
 Db 15430 GKYIILTENSOGSKSAFVNVRLDTPGPPQDLKVKEVTKTSVTLTWDPPLDGGSKIKNY 15489
 Qy 108 ----- 107
 Db 15490 IVEKRESTRKAYSTVATNCHKTSMKVQDQEGCSYVFRVLAENEYIGLPAETAESVKAS 15549
 Qy 108 ----- 107
 Db 15550 ERPLPPGKITLMDVTRNSVLSWEKPEHGGSRILGIYIVEMOTKSGDKWATCATVKVTEA 15609
 Qy 108 ----- 107
 Db 15610 TITGLIOGEYSFRVSAQNEKISDPRQLSVPIAKDLVIPPAPKLLFNTFTVLAGEDLK 15669
 Qy 108 ----- 107
 Db 15670 VDVPPFGRPTPAVTHKONVPLKQTRVNAESTENNSLLTIKDACREDVGHYVVKLTNSA 15729
 Qy 108 ----- 107
 Db 15730 GEAIETLNVILDKPGPPTGVPVKMDEVTDASITLSWGPYPKYGGSINNYIVEKRDSTT 15789
 Qy 108 ----- 107
 Db 15790 TWOIVSATVARTTIKACRLKTGCEYQFRIAENRYKSTYLNSEPTVAQYPPFKVPGPGT 15849
 Qy 108 ----- 107
 Db 15850 PVVTLSSRDSMEVQWNEPISDGGSRVIGYHLERKERNISILWVKLNKTPIPQTKFTTGLE 15909
 Qy 108 ----- 107
 Db 15910 EGVEYEFVSAENIVGIGKPSKVSECVARDPCDPPGRPEAIIVTRNSVTLQWKKPTYDG 15969
 Qy 108 ----- 107
 Db 15970 GSKITGYIVEKKELPEGRWMAKSFNTIIDTHFEVTGLVEDHRYEFVRVIAENAGVFPSEPS 16029
 Qy 108 ----- 107
 Db 16030 ESTGAITARDEVDPPIRISMDPKYKDTIVVHAGESFKVDADIYKGIPTIQIWKQDQELSN 16089

Qy	108	-----	107
Db	16090	TARLEIKSTDATSLSVKDAVRVDSGNVILKAKNVAGERSVTNVKVLDRPPEGPVVI	16149
Qy	108	-----	107
Db	16150	SGVTAEXCTLAWKPLQDGGSDIINYIVERRETSRLVMTVVVDANVOTLSCKVTKLLEGNE	16209
Qy	108	-----	107
Db	16210	YTFRIMAVNKYGVEPLESEBPVAKNPFVVPDAPKAPEVITVTKDSMIVWERPASDGG	16269
Qy	108	-----	107
Db	16270	EILGVLEKRDKEGIRWTRCHKRLIGELRLRVLTGLIENHDYEFVSAENAAGLSEPPS	16329
Qy	108	-----	107
Db	16330	AYQACDPIYKPGPPNPKVIDITRSSVFLSWSPKIYDGGCEIOGYIVEKCDVSGEWTM	16389
Qy	108	-----DL 109	
Db	16390	CTPPTGINKTNIYEKLEKHEYNFRICAINKAGVGEHADVPGPPIVEEKLEAPDIDL	16449
Qy	110	EL-----	111
Db	16450	ELRKIINIRAGSLRLFVPKIGRPTPEVKWGVDEIRDAIIDVTSSFTSLVDNVNRY	16509
Qy	112	-----	111
Db	16510	DSGKYTLTLENSGTSKSAFVTVRVLDTPSPVNLKVTETKDSVITWEPPLDGGSKIK	16569
Qy	112	-----	111
Db	16570	NVIVEKREATRKSAAVNTCHKNWKIDQLEGCSYFRVTAENEYIGLPAQTADPIK	16629
Qy	112	-----	111
Db	16630	VAEVQPPGKITVDVTRNSVLSWTKPEHDGSKIIOYIVEMQAKHSEKSECARVKS	16689
Qy	112	-----	111
Db	16690	QAVITNLTOGEYLFVRVAVNEKGRSDPRSLAVPIVAKDLVIEPDVKPAFSSYSVQVGD	16749
Qy	112	-----	111
Db	16750	LKIEVPISGRPKTTITWKDGLPLKQTRINVTDSLDTLTLSEIKETHKDDGGQGITVAN	16809
Qy	112	-----	111
Db	16810	VVGQKTASIEIVTLDKPDPKGVKFDVSAESITLSWNPPLYTGCCQITNIVOKRDTT	16869
Qy	112	-----	111
Db	16870	TTVWDVVSATVARTTLKVTKLKTGTEYQFRIFAENRYGOSFALESPIVAQVPYKEPGPP	16929
Qy	112	-----	111
Db	16930	GTPFATAISKDSMVIQWHEPVNNGSPVIGYHLERKERNISILWTKVNTKTIIHDTQFAON	16989
Qy	112	-----	111
Db	16990	LEEGIEYFRVYAENIVGVKASKNSECYVARPCDPPGTPPEIMVKRNEITLOWTKPVY	17049
Qy	112	-----	111
Db	17050	DGGSMTGYIVEKRDLPDGRMKASFTTNVETQFTVSGLTEDQRYEYFRVIAKNAAGAISK	17109
Qy	112	-----	111
Db	17110	PSDSTGPIAKDEVELPRISMOPKFRDTIVVNAGETFRLEADVHGKPLPTIEWLRGKEI	17169
Qy	112	-----PDGNR 116	

Db	17170	BESARCEIKNTDFKALLIVKDAIRIDGGQYILRASNVAGSKSPVNVKVLDRPPEGPV	17229
Qy	117	QVRGVT-----OLGG-----	126
Db	17230	QVTGVTSEKSLTSPPLQDGGSDISHYVVEKRETSRLAWTVVASEVVTNSLKYTKLLEG	17289
Qy	127	-----	126
Db	17290	NEYVFRIMAVNKYGVEPLESAPVLMKPPFVLPGPPKSLLEVNTIAKDSMTVCWNRPSDG	17349
Qy	127	-----	126
Db	17350	GSEIIIGYIVEKDRSGIRWIKCNKRITDLRLRVLTGLTEDHEYEFVSAENAAGVCEPSP	17409
Qy	127	-----ACSP-----	130
Db	17410	ATVYKACDPVFKPGPPTNAHIVDTTKNSITLAWGKPIYDGGSEILGYVVEICKADEEEM	17469
Qy	131	-----	130
Db	17470	QIVTPQTLRVTRFEISKLTEHOEYKIRVCALNKVGLGEATSVPGTVKPEDKLEAPDL	17529
Qy	131	-----TWS-----	133
Db	17530	DSELKGIIVRAGGSARIHIHPFKGRPTPEITWSREEGEFTDKVQIEKGVNYTOLSINDCD	17589
Qy	134	-----	133
Db	17590	RNDAGKIILKLENSGSKSAFVTVKVLDTFGPPONLAVKEVKDSAPFLVMEPPIIDGAK	17649
Qy	134	-----	133
Db	17650	VKNYVIDKRESTRKAVANVSSKCKTSFKVENLTEGAIYFVRVMAENEFGVGVVETVDA	17709
Qy	134	-----	133
Db	17710	VKAAEPSPGKVTLTDVTSQTSASLMWEKPEHDGGRVLGVYVEMOPKGTAKSVIAESK	17769
Qy	134	-----	133
Db	17770	VCNAVWTGLSSGEYQPRVKAYNEKGS DPRVLGVPVIAKDLTIQPSLKLPTNTYSIQAG	17829
Qy	134	-----	133
Db	17830	EDLKIEIPVIGRPRPNISWVKGEBPLKQTRVNVETATSTVLHIKEGNKDDFKYTVTA	17889
Qy	134	-----	133
Db	17890	TNSAGTATENLSVILEKPGPPVGPVRPDEVSADEFVVISWEPAYTGCQISNVIVEKRD	17949
Qy	134	-----	133
Db	17950	TTTTWHMVSATVARTTIKITKLKTGTEYQFRIFAENRYGKSAPLDSKAVIVQYPFKEPG	18009
Qy	134	-----	133
Db	18010	PPGTPFTVISKQMLQVWHEPVNDGGTKIIGYHLEOKEKNSILWLKNTKPTIQTKEPKT	18069
Qy	134	-----	133
Db	18070	TGLEGLEEYEPKVSANENVIGIKPSKVSECFVARPCDPPCPPEAIVITRNNVTLKWKPK	18129
Qy	134	-----	133
Db	18130	AYDGSKITGYIVEKKDLPDGRMKASFTNVLETEFTVSGLVEDQRYEYFRVIAARNAAGNF	18189
Qy	134	-----	133
Db	18190	SEPSDSSGAITARDEIDAPNASLDPKYKDVIVVHAGETFRVLEADIRGKPIPDVWWSKOKG	18249
Qy	134	-----	133

Db 18250 ELEETAARMEIKSTIQKTTLLVVKDCIRTDGQYILKLSNVGTSKSIPIITVKVLD RCPGPE 18309
Qy 134 ----- 133
Db 18310 GPLKVTGTAECYLA WNPPLQDGGANISHVIEKRETSRLSWTOVSTEQALNYKVTKL 18369
Qy 134 ----- 133
Db 18370 LFGNEYIIFRVMAVNYKIGEPLESPVTACNPKYKPPGPPSTPEVSAITKDSMVVTWARPV 18429
Qy 134 ----- 133
Db 18430 DDGGTEIEGYILEKXDKEGVRWTKCNKKTLLDRLRLVTLTGEGHSYEFRAAENAGVGE 18489
Qy 134 ----- 133
Db 18490 PSEPSVFRACDALYPGPPSPKVTDTSRSSVSLAWSKPIYDGGAPVKGYVVEVKEAAA 18549
Qy 134 ----- 133
Db 18550 DEWTTCTPPTGLOGKQFTVTKL KENTENFRICAINSEGVGEPATLPGSVVAQERIEPPE 18609
Qy 134 ----- 133
Db 18610 IELDADLRKVVVLRASATLRLFTVIKGRPEPEVKWEKABGILTDRAQIEVTSFTMLVID 18669
Qy 134 ----- 133
Db 18670 NVTRFDGRYNLTLNENSGSKTAFVNVRLDPSAPVNLTIREVKKDSVTLSEWEPPLIDG 18729
Qy 134 ----- 133
Db 18730 GAKITNYIVEKRETTKAYATITNNCTKTTFRIENLQEGCSYFFRVLASNEYIGLPAET 18789
Qy 134 ----- 133
Db 18790 TEPVKYSEPLPPGRVTLVDVTRNTATIKWEKPESDGSKITGYVVMOTKGSEKSTCT 18849
Qy 134 ----- 133
Db 18850 QVKTLEATISGLTAGEEYFVRAAVNEKGRSDPROLGPVPIARDIEIKPSVELPFHTFNV 18909
Qy 134 ----- 133
Db 18910 KAREQLKIDVPFKGRPOATVNRKDGQTLKETTRVNVSSKVTLSLSIKEASKEDVGTYE 18969
Qy 134 ----- 133
Db 18970 LCVSNSAGSITVPITIIIVLDRPGPPGPIRIDEVSCDSITISMNPPEYDGGCQISNYIVEK 19029
Qy 134 ----- 133
Db 19030 KETTSTTHIVSOAVARTSIKIVRLTTGSEYQFRVCAENRYGKSSYSESSAVVAEYFPSP 19089
Qy 134 ----- 133
Db 19090 PGPGTPKVYHATKSTMLVTWQVPVNDGGSRVIGYHLEYKERSILWSKANKILIADTQM 19149
Qy 134 ----- 133
Db 19150 KVSGLDEGLMEYRVAENIAGIGKSKCEPVAPDCDPGQPEVNTITRKSLSLWS 19209
Qy 134 ----- 133
Db 19210 KPHYDGGAKITGYIVERRELDPGRWLKCNNTYIQETIYFEVTELTEDORYEFVRFARNAAD 19269
Qy 134 ----- 133
Db 19270 SVSEPESTGPIIVKDDVEPRVNMVDFRDVIVVKAGEVLKINADIAGRPLPVISWAKD 19329
Qy 134 ----- 133
Db 19330 GIEIEERARTEIISTDNHTLLTVKDCIRRDGTQYVTLKQNVAGTRSVAVNCKVLDKPGPP 19389

Qy 134 ----- 133
Db 19390 AGPLEINGLTAECSLSWGRPOBDGGADIDYIIVEKRETSRLAWTICEGELOMTSCKVTX 19449
Qy 134 ----- 133
Db 19450 LLKONEYIFRVTVGNKYGVEPLESVAIKALDPFTVPSPPTSLEITSVTKESMTLCHSRP 19509
Qy 134 ----- 133
Db 19510 ESDGGSEISGYIIERREKNSLRVVRVNNKPVYDLRVKSTGLREGCEYEYRVAENAAGLS 19569
Qy 134 ----- 133
Db 19570 LPSETSPILRAEDPVFLPSPSPKPIVDSGKTTITIAVWKPLFDGGAPITGYTVVEYKXSD 19629
Qy 134 ----- 133
Db 19630 DTDWKTISIQLRGTEYTISGLTTGAEYVFRVKS VNKVGASDPDSSDPQIAKEREEBPLF 19689
Qy 134 ----- 133
Db 19690 DIDSEMRKTLIVKAGASFTMTVPGRGPVNVLWSKPD TDLRTRAYVDTTDSRTSLTIEN 19749
Qy 134 ----- 133
Db 19750 ANRDSGKYTLTIQNVLSAASLTIVVKVLDTPGPPTNITVODVTKESAVLSWDVPENDGG 19809
Qy 134 ----- 133
Db 19810 APVKNYHIEKREASKAWSVTNNCNRLSYKVTNLQEGAIYYFRVSGENEGVGI PAETK 19869
Qy 134 ----- 133
Db 19870 EGVKITEKSPPEKLGVTISISKDSVSLTLWKPHEHGSRI VHYVVEALEKGQKNVVKAV 19929
Qy 134 ----- 133
Db 19930 AKSTHHVVGLRENSYFFRVFAENQAGLSDPRELLLPVLIKEOLEPEIDMKNPSPSHV 19989
Qy 134 ----- 133
Db 19990 YVRAGSNLKVDPISGKPLPKVTLSDRGVPLKATMRFNTEITAENLTLNKESVTADAGR 20049
Qy 134 ----- 133
Db 20050 YEITAANSSTTKAFINIVVLDRPGPPTGPVVISDITEESVTLKWEPPKYDGGSOVTNYI 20109
Qy 134 ----- 133
Db 20110 LLKRETSTAVMTEVSATVARTMMKMKLTGTGEYQFRIKAENRFGISDHIDSACVTVKLP 20169
Qy 134 ----- 133
Db 20170 YTTGPPSPFTWVNTRESITVGMHEPVNSGSAVVG YHLEMKDRNSILWOKANKLVIRT 20229
Qy 134 ----- 133
Db 20230 THFKVTIISAGLIYEFVRYAENAGVKPSHPSEPLAIDACEPPRNVRITDISKNSVSL 20289
Qy 134 ----- 133
Db 20290 SMOQPAFDGSKITGYIVERDLDPGRWTKASFTNVTETOFIISGLTQNSQYEFVRFARN 20349
Qy 134 ----- 133
Db 20350 AVGSINPSEVGPVITCIDSYGGVIDLPLEYEVVKYRAGTSVKLRAGISGKPAPTIEW 20409
Qy 134 ----- 133
Db 20410 YKDDKELOTNALVCVENTTDLASILIKOADRNLNSGCEYLKLRNAMGSASATIRVOILDKP 20469

QY 134 -----CLIT 137
Db 20470 GPPGGIEFKVTABKITLLWRPPADGGAKITHYIIVEKRETSRVVMSVSEHLEECIIT 20529
QY 138 ----- 137
Db 20530 TTKIIGNEYIFRVRAVKNYKIGEPLESDSVAKNAFVTPGPGIPEVTKITNSMTVVW 20589
QY 138 ----- 137
Db 20590 SRPIADGSDISGYFLEKRDKSLGFWKVLKETIROTKVTLGTLTENSQYRVCAVMAA 20649
QY 138 ----- 137
Db 20650 GOGPFSEPFYKAADPIDPPGPAKIRIADSTKSSITLGSXPKVYDGGSAVTGYVVEIR 20709
QY 138 ----- 137
Db 20710 QGEEBETVSTKGEVTTTEYVVSNLKPGVNYFRVSAVNCAGOGEPENNEPVQAKDIL 20769
QY 138 ----- 137
Db 20770 EAPEIDLVALRTSVIAKAGEDVQLIPFKGRPPPTVTRKDEKILGSDARYSIENTDSS 20829
QY 138 ----- 137
Db 20830 SLLTIPQVTRNDTGKYLITENGVEPKSSTVSXKVLDTPAACQKLOVKHVSRGVTLLW 20889
QY 138 ----- 137
Db 20890 DPPLIDGSPINIYVIEKRDATKRTWSVSHKCSSTSKLIDLSEKTPPPFRVLAENEIG 20949
QY 138 ----- 137
Db 20950 IGEPCETTEPVKAAEVPAPIRDLMSKDSKTSVILSWTKPFDGSGVITEYVVERKGE 21009
QY 138 ----- 137
Db 21010 QTWSHAGISKTCETEVSQLEQSVLEPRVFAKNEKGLSDPVTGPTVVKELIITPEVDLS 21069
QY 138 ----- 137
Db 21070 DIPQAVTVRIGHNVHLELPYKGPSPISWLKDGLPLKESEFVRFSKTENKITLSIKNA 21129
QY 138 ----- 137
Db 21130 KKEHGGKYTVILDNAVCRIVAITVITLGPSPKPGPIRFEDEIKADSVILSWDPEDNG 21189
QY 138 ----- 137
Db 21190 GEITCYSIEKRETSQTNWKMVCSSVARTTFKVPNLVKDAEYQFRVRAENRYGVSQPLVSS 21249
QY 138 ----- 137
Db 21250 IIVAKHQFRIIPGPGKPVYINVTSDGMSLTWDAPVYDGGSEVTGFHVEKKERNISILQKV 21309
QY 138 ----- 137
Db 21310 NTSPTISGREYRATGLVEGLDYQFRVYAENSAGLSPPDSKFTLAVSPVDPGPDYDV 21369
QY 138 ----- 137
Db 21370 TRETITLKNPPLRDGSKIYGVYSIEKQGNERNWRCNFTDVSECOYTVTGLSPGDRYEF 21429
QY 138 ----- 137
Db 21430 RIIRNAVGTISPPSQSGIIMTRDENVPPIVEFGPEYFDGLIINKGESLRKALVQGRP 21489
QY 138 ----- 137
Db 21490 VPRVTFKDGVEIEKRMNMEITDVLGSLFVRDATRHRGVYTVAKNAGSAGAKAEIKV 21549
QY 138 ----- 137

Db 21550 KVQDTPGVVGPPIRFTNITGEKMTLWMDAPLNDGCAPIITHYIEKRETSRLAWALIEDKC 21609
QY 138 ----- 137
Db 21610 EAQSYTAIKLINGNEYQFRVSAVKNKFGVGRPLDSDPPVVAIQIYTVDPAPGIPESNITGN 21669
QY 138 ----- 137
Db 21670 SITLTWAPESDGGSEIQOYLIERREKSTRVWVVISKRPISETRFKVGTGLTEGNEYEFH 21729
QY 138 ----- 137
Db 21730 VMAENAAVGVPASGISRLIKCREPVNPGPPTVVKVTDTSKTTVSLEWSPKVPDGMELI 21789
QY 138 ----- 137
Db 21790 GYIIEMCKADLGWHKVNAAEACVKTTRYTVTLQAGEEYKFRVSAINGAGKGDSCVETGTI 21849
QY 138 ----- 137
Db 21850 KAVDLTAPELDIDANFKQTHVVRAGASIRLFIAYQGRPTTAVMSKPDNSLSLRADIHT 21909
QY 138 ----- 137
Db 21910 TDSPTLTVENCNRNDAGKYTLTVENNSGSKSITFTVKVLDTPGPPGPIPKDVTRGSAT 21969
QY 138 ----- 137
Db 21970 LMDAPLDGGARIHYYVVEKREASRRSQVISEKCTQIIFKVNDLAEGVPYFRVSAVN 22029
QY 138 ----- 137
Db 22030 EYGVCEPYEMPEPIVATEQPAPRRLDVDTSSSAVLAKLPDHDGGSRTGYLLEMRQ 22089
QY 138 ----- 137
Db 22090 KGSDFWEAGHTKOLTFTVERLVEKTEYEFVKAKNDAGYSEPREAFSIIKEPOIEPT 22149
QY 138 ----- 137
Db 22150 ADLTGITNOLITCKAGSPFTIDVPISGRPAKVTKLEMLKEDTRVSIITTKDRTTLT 22209
QY 138 ----- 137
Db 22210 VKDSMRGDSGRYFLTLENTAGVKFTSVTVWIGRPGVPTGPIEVSSVSAESCVLSWGPBK 22269
QY 138 ----- 137
Db 22270 DGGGTEITNYIVEKRESGTTAMQLVNSSVYKRTQIKVTHLTMYEYSFRVSGENRFGVSKP 22329
QY 138 ----- 137
Db 22330 LESAPIIAEHPFPVPPSPAPTRPEVYHVSAVANSIRWEEPYHDGSKTIIGYVVEKERNITIL 22389
QY 138 ----- 137
Db 22390 WVKENKVPCECNKYVTGLVEGLEYOFRTYALNAAAGVSKASEASRPFMAQNPVDAPGRPE 22449
QY 138 ----- 137
Db 22450 VTDVTRSTVSLIWSAPAYDGGSKVGYIIERKPVSEVGGGRWLKCNVYTIIVSDNFTVTAL 22509
QY 138 ----- 137
Db 22510 SEGDTYEFVRLAKNAAGVISKSESTGPTVTCRDEYAPPAKELARLHLDLVTIRAGSDLV 22569
QY 138 ----- 137
Db 22570 LDAAGVKPEPKIITWKDDELDCESKVSLOYTGKATAVIKFCDRSDSGKYTLTVKNAS 22629
QY 138 ----- 137

Db	22630	GTKAVSMVKVLDSPGCKLTVSRVTOEKTCLAWSLQEDGGABIITHYIVERRETSRLN	22689
Qy	138	-----	137
Db	22690	WVIVEGECPTLSYVTVRLIKNNYIFRVRANKYGGVPVSEPIVARNSETIPSPGP	22749
Qy	138	-----	137
Db	22750	EEVGTGKEHIIQWTKPESDGGNEISNYLVDRKESLSRLWRTRVNDYVYDTRLKVTSLM	22809
Qy	138	-----	137
Db	22810	EGCDYQFRVTAVNAGNSEPSEASNFI SCREPSYTPGPSAPRVVDTTKHSLAWTKPM	22869
Qy	138	-----	137
Db	22870	YDGGTDIVGVLEMOEKOTDQWYRVHTNATIRNTEFTVDDLKMGOKYSFRVAANVKGMS	22929
Qy	138	-----	137
Db	22930	EYSESTAEIEPVERIEIPDELAADDLKKTVTIRAGASLRMLVSVSGRPPPVITWSKQIGID	22989
Qy	138	-----	137
Db	22990	LASRAIIDTTESYSLLI VDKWRYDAGKYTIEAENOSGKSATVLKVYDTPGPCPSVKV	23049
Qy	138	-----	137
Db	23050	KEVRSDSVTITWEIPTIDGAPVNNYIVEKREAAAMFAKTVTTKCSKTLYRISGLVEGTM	23109
Qy	138	-----	137
Db	23110	YFPRVLPENIYGIGECETSDAVLVSEVPLVPAKLEVVVDTKSTVLAWEKPLYDGGSL	23169
Qy	138	-----	137
Db	23170	TGVVLEACKAGTERMKVWTLKPTVLEHTVTSLNEGEQYLFRITRAQNEKGVSEPRETVTA	23229
Qy	138	-----	137
Db	23230	VTQDLRVLPITDLSTMPQKTHVPAGRPVELVPIAGRPPPAASWFFAGSKLRESERV	23289
Qy	138	-----	141
Db	23290	VEHTKVAKLTIRETIRDTGEYLELKNVTGTSETIKVILDKPGPTGPIKIDEIDA	23349
Qy	142	-----	141
Db	23350	TSITISWEPPELDGAPLSGVVVEORDAHRPGWLPSVESVTRSTFKFTRLTEGNEYVFRV	23409
Qy	142	-----	141
Db	23410	AATNRFGISYLOSEVIECRSSIRIPGPPTLQIFDVSROGMLTWYPPEDDGGSQVTGY	23469
Qy	142	-----	141
Db	23470	IVERKEVRADRWVRVNVKVPVMTYRSTGLTEGLEYEHRVTAINARGSGKPSRSPKPIVA	23529
Qy	142	-----	141
Db	23530	MDPIAPPGKQNPVTDTRTSVSLAWSVPDEGGSKVTGYLIEMQKVDOHEWTKCNTTP	23589
Qy	142	-----	141
Db	23590	TKIREYTLTHLPOGAERYFRVLACNAGGPEAEVPGTVKVTMELYPDYELDERVOEGI	23649
Qy	142	-----	141
Db	23650	FVROGGVIRLTIPIKGRFPPIKWTKEQDISKRAMIATSETHELVIKEADRGDSCTYD	23709
Qy	142	-----	141
Db	23710	LVLENKCGKAVYIKVRVIGSPNSPEGPLEYDDIQVRSVRVSWRPPADGGADILGVILE	23769
Qy	142	-----	141
Db	23770	RREVPKAAWYTIIDSRVGTSLVVKGLKENVEYHFRVSAENQFGISKPLKSEEPVTPKTPL	23829
Qy	142	-----	141
Db	23830	NPPEPPNPPEVLDTVKSSVLSWSRPPKDDGGRVTGYIYIERKETSTDKWVRHNKTOITT	23889
Qy	142	-----	141
Db	23890	TMYTVTGLVPDAEYQFRI IAQNDVGLSETSPASEPVVCKDPFOKQSQPQGBELEILSISKDS	23949
Qy	142	-----	141
Db	23950	VTLOWEKPECDGGKEILGYWVEYRQSCDSAWKSKNKERI KDKOPTIGGLLEATEYERVP	24009
Qy	142	-----	141
Db	24010	AENETGUSRPRRTAMSIKTKLTSGEAPGIRKEMKDVTTKLGEAAQLSCQIVGRPLPDIKW	24069
Qy	142	-----	141
Db	24070	YRFGKELIOSRKYKSSDGRTHTLTVMTEEQDEGVVTCIATNEVGEVETSSKULLQATP	24129
Qy	142	-----	141
Db	24130	QFHPGYPLKEYYGAVGSTLRLHVYIIGRPVPAMTWPHGOKLQNSENITIENTEHTHYTHL	24189
Qy	142	-----	141
Db	24190	VKNVQRKTHAGKYKVQVLSNVFGTVDAILDVEIODKPKPTGPIVIEALLKNSAVISWKP	24249
Qy	142	-----	141
Db	24250	PADGGSWITNVVVEKCEKEGAEWQLVSSAISVTTCRIVNLTENAGYFRVSAQNTFGI	24309
Qy	142	-----	141
Db	24310	SDPLEVSSVWVIKSPFEKPGAPGPTITAVTKDCSCVAVWPKPPADGGGAKIRNYILEKREK	24369
Qy	142	-----	141
Db	24370	KONKWI SVTTEIETRETVSVKNLIEGLEYEFRVKCENLGSESEISEPIPTKSDVDPIQ	24429
Qy	142	-----	141
Db	24430	APHFKEELNLRVYQSNATLVCKVTHGHPKPIVKWYRQGEKIIADGLKYRIOEFKGGYHQ	24489
Qy	142	-----	141
Db	24490	LIIASVTDATVYQVRATNCGSVSGTASLEVEVPAKIHLPKLTUEGMVHALRGEVVS	24549
Qy	142	-----	141
Db	24550	IKIPSGKPPDVITWQKQDLIDNNGHYQVIVTRSFVSLVPNGVERKDAGFYVVCANUR	24609
Qy	142	-----	141
Db	24610	FGIDQKTVLVDVADVPDPPRGVKVSDVSRDSVNLTWTEBPASDGGSKITNYIIVEKCATTA	24669
Qy	142	-----	141
Db	24670	RMLRVQARETRYTVINLFGKTSYQFRVIAENKFGLSKSPSESEPTITKEDKTRAMNYDE	24729
Qy	142	-----	141
Db	24730	EVDRETREVMTKASHSTKELYKYMIAEDLGRGEFGIVHRCVETSSKTKYMAKFVKVKG	24789
Qy	142	-----	141
Db	24790	TDQVLVKKEISILNIARHRNLIHLHESFESMEELVMIPEFISGLDIFERINTSAFELNBR	24849
Qy	142	-----	141

```
QY 142 -----FDL----- 144
Db 24850 EIVSYHVCEALQFLSHNIGHFDIRPENIIYOTRRSSTIKIEFCQARQLKPGDNFRL 24909
QY 145 ----- 144
Db 24910 LFTAPEYYAEVHOHDVVSTATDMWSLGLTVYVLLSGINPLAETNQOIIENIMNAEYTF 24969
QY 145 ----- 144
Db 24970 DEEAFKEISIEAMDFVDRLLVKERKSRMTASEALQHPWLKOKIERVSTKVIRTLKHRRYY 25029
QY 145 ----- 144
Db 25030 HTLIKDLNMVVSAAISCGGAIRSQKGVSAKVKVASIEIGPVSGOIMHAVGEGGHVK 25089
QY 145 ----- 144
Db 25090 YVCKIENYDQSTQVTWYFGVQLENSEKYEITYEDGVAILYVKDITKLDGTYRCKVVND 25149
QY 145 ----- 144
Db 25150 YGEDSSVAELFVKGVREYVYCYRRTMKKIKRRTDTMRLLRPPFTLPLYNKYAYGEN 25209
QY 145 ---GVTI 148
Db 25210 VRFGVTI 25216

RESULT 5
Q9EWA1 PRELIMINARY; PRT: 9507 AA.
ID Q9EWA1
AC Q9EWA1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE PimS2 protein.
GN PIMS2.
OS Streptomyces natalensis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=68242;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20547809; PubMed=11094342;
RA Aparicio J.F.; Fouces R.; Mendes M.V.; Olivera N.; Martin J.F.;
RT "A complex multienzyme system encoded by five polyketide synthase
RT genes is involved in the biosynthesis of the 26-membered polyene
RT macrolide pimarinin in Streptomyces natalensis.";
RL Chem. Biol. 7,895-905(2000).
DR EMBL; AJ278573; CAC20921.1; -.
DR HSSP; P25715; IMLA.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR002198; ADH short.
DR InterPro; IPR004410; Fabb.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000794; ketoacyl-synt.
DR InterPro; IPR001005; Myp_DNA_binding.
DR InterPro; IPR003880; Ppantne_attach.
DR Pfam; PF00698; Acyl_transf; 6.
DR Pfam; PF00106; adh_short; 3.
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt_C; 6.
DR Pfam; PF00550; pp-binding; 6.
DR TIGRfams; TIGR00128; fabb; 6.
DR PROSITE; PS50075; ACP_DOMAIN; 6.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00037; MYB_I; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 5.
KW Phosphopantetheine.
SQ SEQUENCE 9507 AA; 994228 MW; 57BD80C8AD37E89F CRC64;
```

```
Query Match 34.2%; Score 264; DB 2; Length 9507;
Best Local Similarity 1.2%; Pred.No. 1.1;
Matches 98; Conservative 19; Mismatches 30; Indels 7866; Gaps 28;
QY 2 AGGILH-----LELLV----- 12
Db 86 SGGFLHDAAEFADSPFGISPREALAMDPOORLLESSHEAFERAGIDPNAVRGTRTGMFV 145
QY 13 -----AVGP----- 16
Db 146 GAMPQERYVPDDDVQGFALTGTTTSVISGRLAYFFGAUGPAVTVDTACSSSLVALHLAA 205
QY 17 ----- 16
Db 206 HSLRQECSLAALAAAGVTVMSPPTTFVFNROGGLSADGRCRSFADSADGTGSEGVLV 265
QY 17 -----DV 18
Db 266 LERLSEARRNGHEILAVIRGSVNOGASNGLTAPNGPSORRVIEQALYSARLSADEVDV 325
QY 19 FQAH----- 22
Db 326 VEAHGTTGLGDPVEAQAALLATVQGRDDDRPLLGSVKSNLGHTQAAAAGMAGVIKMWLA 385
QY 23 ----- 22
Db 386 MRHGVLPTLHVDAPSSHVDWTQGAVALRLTEHVMPQSGOPRAGVSFGLSCTNAHTIL 445
QY 23 -----QEDT----- 26
Db 446 EEAPAAPSPGAGQEDTAPVAEPLAAGAVPWLVSGRTRREALRAQAARLLEHLTSRPL 505
QY 27 ----- 26
Db 506 DVAYSLATARSGLHRAAFTTAAGDCARAALTALADGTPAPGLVQDTARTSKLAFAG 565
QY 27 ----- 26
Db 566 QGSORPQMGRELAAFPVFATALDEVLGHFDEGLERPLKDVLPAAEGTPEAALLDQTGYA 625
QY 27 ----- 26
Db 626 QPALFAIEVALYRLAESFGITPDLFAGHSIGEIAAAHVAGVFTLADAAALVLARGRLMQA 685
QY 27 ----- 26
Db 686 LPEGAMV/SLEATEEEVLPLEDRPDKLSIAAVNGPAAVVIAGEADDVTVAEHFDA LGR 745
QY 27 ----- 26
Db 746 RTKRLRVSHAFSPLMPEMLAEFRFVAVSLRTPQAPVTPVVSGLTGLTATVEQLTSPDYWL 805
QY 27 ----- 26
Db 806 DHARHAVRFADALRLARRRRRLPRMGPDGLSAMAQSCCRGHGRRRRRLRHARRVP 865
QY 27 ----- 26
Db 866 GRPEAETLTAALAGLHTRGVAVQWEPYFQGTGARRADLPTYAFQRRRYWPKSLPATGGDV 925
QY 27 ----- 26
Db 926 RAAGLGAHHPLTAAVSVANSOGLLTLGRLSRRTHPWLADHAVRGTVLLPGTAFLELAV 985
QY 27 ----- 26
Db 986 RAGDEACGRVEELTAAPLVLPPEGGVQVQVWVSPDASGRRAVSVHSRDPGPEELPWT 1045
QY 27 ----- 26
Db 1046 QHAAAGTLVAGEHHAGFDATVMPPADATPLDLDGFYARMADTDFGFGYPLFOGLRAARSGD 1105
QY 27 ----- 26
```

```
Db 1106 DVYAEVALPDSGNASTESSAAGLHPALLDAALHAAAFVDLGEDARGGLPFSQDVTLLH 1165
QY 27 ----- 26
Db 1166 ASGASTVRVKLTDPGDDAVALAVADTTGAPVASIGSLVLRTPDEQIDAAHSLVRDALFE 1225
QY 27 ----- 26
Db 1226 LRNSEARQTAHEAPASAAVLGPDLPGLERLALPTDLTVTAVPGPDAAHPADPVVLAPVAG 1285
QY 27 ----- 34
Db 1286 TTEDDVTSVHARTAHVLERLRTWLTDERYAASRLVFVTRGALATDARQASDPVAAAAMVG 1345
QY 35 ----- 34
Db 1346 LVRAAQSEHPGRFGLDLDPTTTEPATEPLLRALALSSDEPQTAVRGTAVLRTARLARAQA 1405
QY 35 ----- 34
Db 1406 PQAAATEWDPECTVVLTGATGGLGRVLARHLVAERGVHRHLLLVSRRGADADGAGELMAELA 1465
QY 35 ----- 34
Db 1466 AHGAQVTLAACDLADRAAVRLLSGVPAEHPVTAVVHSAGSLDDGVIASLTPERISAVLR 1525
QY 35 ----- 34
Db 1526 PKADAVWLHEATRDLDLAFWPFSSLSGTVGAAAGQNYAANAFLDALAQLRRAOGLPG 1585
QY 35 ----- 34
Db 1586 LSLANGPWAPTAGTGGTDDDLRLARMGTTPALTEEGIALFDAATATDAVLLPTRID 1645
QY 35 ----- 34
Db 1646 LSVLRVQSEIPLMRGLRTPARRSVVSASSETAVTLVQOLSRLAETDRREVVLDLVAGV 1705
QY 35 ----- 34
Db 1706 AAVLGHAGTADIDPRRPLRELGFDSLTAVELNRLSAATGURTAATVIFDHTVDALAAH 1765
QY 35 ----- 34
Db 1766 LLDGLMGSEAAAAADAPRTAADDDPIVIGMSCRYPGGVASPEDLWRLVSEGTDASGLP 1825
QY 35 ----- 34
Db 1826 TDRGWLDALYDPDPDRTGTSYRFGFLHTAADFPAPFGMSPREALATDSQQORLLLEA 1885
QY 35 ----- 34
Db 1886 SCEAVERAGIDPVSLSRSGTGVFAGVMYNDYATVLSGOQFEGHOOGTAPSVASGRISYT 1945
QY 35 ----- 34
Db 1946 LGLEGPVTVDTACSSSLVAMHWAMQALRAGECSLALAGGVTVMTSTPGALIEFSRQRLS 2005
QY 35 ----- 34
Db 2006 PDGRCKAFSDGADGVGWSEGVGLVLERLSDARRNGHRVLAVVRSVAVNQDGASNGLTAP 2065
QY 35 ----- 34
Db 2066 NGPAQORVIRQALSAGLSATDIDAVEGHGTGTTGLDPIEAQALLAAYGQDREAGSPLLL 2125
QY 35 ----- 37
Db 2126 GSVKSNIGHTQAAAGVAGVIKVMAMRHCVLPRTLHADEPSSHVEWDGAVRLLTTEETPW 2185
QY 38 ----- 37
```

```
Db 2186 POADRPRRAAVSSFGSGTNAHLIVVEEPAVAERREEPALTPAVVPWTLGSKSRAALRDO 2245
QY 38 ----- 42
Db 2246 AARLSFLEERPALDPADVALSLATRSADFQORAVVTGDRDELLRALADLAADRPGPALT 2305
QY 43 ----- 46
Db 2306 EGEVGGAGKLAUVFSGQSORPGAGRELAARPPVFAQALDEVTAALDPLHDLRPLKILPA 2365
QY 47 ----- 46
Db 2366 PEGSPEALLDRTEWTQPALFAIGVALHRLLTGTWGRIPDVLGHSIGEITAAHVAGVLSL 2425
QY 47 ----- 46
Db 2426 PDAARLVIARGRLMQALPAGGAMISLEATEDETAPILLAGRDHEVSLAAVNGPRSVVVAGE 2485
QY 47 ----- 46
Db 2486 LAAAEIEAAHFADRGRTKRLTVSHAFSPLMEPALDGLREAAAEILTYHAPDVTVISGLT 2545
QY 47 ----- 46
Db 2546 GRPATEAELRSADYVYVHARGAVRFADTLURAAHDCGAGTFLSLGPDOSVLSAAAQDALGDD 2605
QY 47 ----- 46
Db 2606 AEAHIVPLLRHGGERCAATALARLHVRTTADMPGYLAGTGARTIDLPTVAFQHERYW 2665
QY 47 ----- 46
Db 2666 PEQAPARLAQGAADPADAEELGWAVEGDATELAALLGLRDEQHASLYALLPALSSWROHR 2725
QY 47 ----- 50
Db 2726 QEKALLDSTRYVAVRPVQATAAPVLDGTWLLVTADGIDSDEILDALRGHGAOFETLVLD 2785
QY 51 ----- 54
Db 2786 DACRDAHLAGRLAAVDDQDRPLTGVLSLLPLADRSGAEGAGAGAGTGAQPSGLALS 2845
QY 55 ----- 54
Db 2846 VLIQALADTGLAAPLWTLTRGAVTTGADDPDLTHPLOAAVMGLGRVALEHPQLMSGILDL 2905
QY 55 ----- 54
Db 2906 PAALDAPATOHLVLSALAAEGGEDQIAVRATGAPGRRLVRHPQAEPLPADAFRTGTVLIT 2965
QY 55 ----- 54
Db 2966 GGTGALGAEAAARWLARSGAEHLLLTSSRGPEAPGAELAAEIEELGARVTLAACDTADRA 3025
QY 55 ----- 54
Db 3026 ALAALLDSLDEHPLTCVWHAAGVQAAALADTPLADAAQMAAKLLGAHLSLLDGH 3085
QY 55 ----- 54
Db 3086 LDPFVLVSSIAGVWGSAGQSAAYSAANNYLDALAEHRAARGLPATSVANGPWAEGMATHE 3145
QY 55 ----- 54
Db 3146 AVTDELHKRGLRFLAPATALAEELRAVVRDVTVTADIDWERYHPVTSTTRPSALFDEL 3205
QY 55 ----- 54
Db 3206 AEVQALIRPEDESAAPEFAARLRGLDQDGOERLLGLDLVRAEAAIALGHDSADATERRAF 3265
QY 55 ----- 54
Db 3266 RDAGFDSLTAVELRKRRLAULTGLALPATLVDPYPTPVALARHLREQLLGTAOESAAPVAV 3325
```


QY	55	-----	54	QY	75	-----	74
Db	3326	PSGGFDEPIAIVGMSCRFPGGVRSRQLWDLVSEGVDAISDFPVNRGWNHGLFHPDPDPN	3385	Db	4406	DOAQILGLGHTAALEHPQRCAGLVDLPELDRRAGERLAAVLTRGSGEDOLAVRSSGVF	4465
QY	55	-----	54	QY	75	-----	74
Db	3306	GTYSTOGFLHDAGEFDSFFGISPREALSMDPOORULLLETTWEAFEHAGIDPTTVHGT	3445	Db	4466	ARRVRAAPADPAPARPWTGRGTTLVGTGTGLAPHLARWLAAQGAHIVLTSRRRGAAP	4525
QY	55	-----	54	QY	75	-----	74
Db	3446	TTGTFIGSTYQBYGLGVEDGSAGHLVGTSPSVLSRGLAYLFLGLEGPATVDTACSSSLV	3505	Db	4526	GAELVQELAEQLCOAEAVACDLTDRDAVQAQMLAALRAEGRTVTVVHTAVTIELAPLDE	4585
QY	55	-----	54	QY	75	-----	74
Db	3506	ALHLACQSLRNGESSLALAGGATVMTNPNPFVAFSRQALAGDRCKAFSEGADGMTLAE	3565	Db	4586	TTLLDDFAKVMDAKVICGARHLDLGGDDDAFVLYSSTAGMWGSAHAAYVAGNAYLNALA	4645
QY	55	-----	54	QY	75	-----	74
Db	3566	GVGLVLERLSDAQORNGHEILAVVRGSAINQDASNGLSAPNGSQQRVIRQALANSGLT	3625	Db	4646	EHRRGARATAVSWGIIWADDLKLGRVDPQGIIRRSGLVFMDPOLALTGLRQALDDEETHL	4705
QY	55	-----	54	QY	75	-----	74
Db	3626	PADIDAVEAHGTGTALGDPIDIAQALLATYCPGRDPDPSPLLLGSKVSNIGHTQSAAGVAGV	3685	Db	4706	AVADVDMGRYYPVPTAARPTLRFEDLPEVTRQLAQAESAETVDSGEFATRLDGLAAVEQDR	4765
QY	55	-----	63	QY	75	-----	74
Db	3686	IKWVLARNQQLPRTLHADTPSSHIDWTPGTVSLLETPDWEGERPRRCVASFSGISGT	3745	Db	4766	LLLDLVRTEAATALGHASPDVLSQRAFRDVGFDLSLTAVDLRNRIRASVTVGLALPSTMVFD	4825
QY	64	-----	63	QY	75	-----	74
Db	3746	NAHTLLEAPRPEATPAGVPVPPAGDAIPWLSARTPGALRAQAAQLAAHLDCGAPDALD	3805	Db	4826	YPNPLALVAFRESLAGSATGTAATTTHTAAADEPIAIIIGMSCRYPGCVGSPEDLWRLVT	4885
QY	64	-----	63	QY	75	-----	74
Db	3806	VGHTLVAARTLFDHRAVVVGTDDASRAALDALATGSGAPGIVQGTADTDGRTVFVFPQ	3865	Db	4886	EGGATGEFPADRCWDAGELYDPPDRAGHTYSTRGGFLHDAADFASFFGISPREALAM	4945
QY	64	-----	67	QY	75	-----	74
Db	3866	GSQWAGMARLLBESPVFAARLTCATALSEFVDSLLDLVRQADGAPTLDRVVDVVPAS	3925	Db	4946	DPQRLILLETSEAWERAGIDPATLRGSGTGTFIGASYQDYAFGAGDPGAEGHLITGTI	5005
QY	68	-----	67	QY	75	SSLLS-----VCG-----	82
Db	3926	FAVMSLAALWTSHTGTPDAVVGHSQGETAAAAVAGALSLEDAAVVVALRSQAIARGLAG	3985	Db	5006	SSVLSGRLSYTYGPEGPAVSLDTACSSSLVALHLACOSLRNGESSLALAGOVSIMSTPGA	5065
QY	68	-----	67	QY	83	-----	82
Db	3986	TGGMLSPLPAADVEQRLAAYEDLSIAAVNGPRSTVVSGATAPLDALQAEVLGVEDIRARR	4045	Db	5066	FVGSFRQAMAADGRCKAYSPOADGMSLAEGVLVLERLSDARRNGHOVLAVIRGSVN	5125
QY	68	-----	67	QY	83	-----	82
Db	4046	IAVDYASHSAQVERVRDELTVLEVPVRPAQVPFFSTVTGDLDTVMDAEYWFNLR	4105	Db	5126	QDGASGLTAPNGPSQQRVIRQALANAEVDPDGIDVIDGHGTGALGDPIDIAQALLATYG	5185
QY	68	-----	67	QY	83	-----	82
Db	4106	TVHFQPAIGELLAQGHFFIEVSSHPVLSMCIQATAEEAGAAAALVGLTRRDGTATDRFL	4165	Db	5186	QGRDPEHPLLGSVKNSIGHTQMASGVASVIMVMMRHAAPRSLHIDRPSHVDWSSG	5245
QY	68	-----	67	QY	83	-----WSQTINP-----EDDTPG-----	96
Db	4166	ASLAEAFVRCADADWSAVFAGTGARRVPLTYAFQREHLWAIPEPSDRPEADPADEFW	4225	Db	5246	AIQLLTEPLPWPATCHPRRAGVSSFGSLGTNVHTILEQAPEDRTBEDTDTAPARPADPTP	5305
QY	68	-----	67	QY	97	-----HADLVL-----	102
Db	4226	TAVEEDVEALSLRLDRTS LAPVLPALSHWRKRRELSTVDSWRYRATWKPLTGLPAA	4285	Db	5306	VPVTLGRTEAGLRAQAGRLTHLADHPLEPLTDLAFSLATSRAALEHRAAVVTTDDPTL	5365
QY	68	-----NITANLT-----	74	QY	103	-----	YIT 105
Db	4286	ALTGTWVVAPEGADDTLTGVLAAHGAGTFTLVLDSDCADRATLTARLTGLDGAADLTG	4345	Db	5366	TRALTALRDATPDGGLTGRDPRGRGLAFLFTCGQSORPCMGRRELYDRHPVYAEALDAVLA	5425
QY	75	-----	74	QY	106	RFDE-----	110
Db	4346	VVSLPADEASATHPTLSRAVTLTVLVALQALGDAGIDAPLWCLUTQGAVSTGRADRLTRP	4405	Db	5426	RFDELDRLRLREILFADPGTPEAELDDTGTQTPALFALEVALFRLAESWGLURPDYVAGH	5485
				QY	111	-----LPDG-----	114

```
Db 5486 SIGELAAAHACVLSLEDACTLVAAAGRLMAALPSGGAMASVEATEDEAAAVLAPYEGRA 5545
Qy 115 -----
Db 5546 AVAAVNTPTSLVSGDEDAVHAHAHFHGLGRRTKQLRVSHAFSPHMDAMDLOFTAAE 5605
Qy 115 -----NRQVRG-----
Db 5606 GIFHAPELPLVSTLTGELILADDLCSAPYWTQVRGTVRFADAVRHLTEHGVTTFFELG 5665
Qy 121 -----
Db 5666 PDAVLGAVRESTGEQTATAVPALRRDRPEAPALTALARLHLHGVVDMDAVPAGGA 5725
Qy 121 -----
Db 5726 RRIDLPTYFORERYWEAANAAPAHADAADAEFWSAVERADLPISLGSSLDLDDTL 5785
Qy 121 -----
Db 5786 TAVPALSSWRKSRERSTVDGWRVYRTTWKPLTGATATAGHPAGTWLVLAPADGDRAWTDA 5845
Qy 121 -----
Db 5846 VAGALCADAVRVEVSTAQRQELAEERLSELAEHGEFAGVLSILLATAGEGAEDADATTG 5905
Qy 121 -----
Db 5906 LLLTATALQALGDAGIDAPLWCVTRTAVADRAEHPARPAQAAVWGLGRVALEHPQRMG 5965
Qy 121 -----
Db 5966 GLVDLPDELGDATLRLAAVLADSGEDQLAVRTTATFVRLAHHRAPAPAAATFRPTG 6025
Qy 121 -----VTQLG-----
Db 6026 TVLVTTGTCALGHHVARWLAEAGAEHLLLVSRRTGDPACAGELAAEIAELGARVTVAACD 6085
Qy 126 -----
Db 6086 TADRALAAVLAAPIDEHPLTAVFHTAGTVDDGTLDTLTPEQFASVLRKVTATRNHET 6145
Qy 126 -----
Db 6146 TREQDLSEFVLFSSVAGTLGAPGQGYAAGNAFLDAFAAHRRAHGLPATSLAWGPWAETG 6205
Qy 126 -----
Db 6206 MAADGTGIQDRVRRGGFTPMPPRLALTALRAIEHDATAALTLDIDWQRYAEVFTATRPS 6265
Qy 126 -----
Db 6266 LFAVDLPQLQVTPAGTPEAALREPALRQLAGLSPAARPRFVLDLVRTOVAVLGHSG 6325
Qy 126 -----
Db 6326 TSAIGERAFSOLGFDLSLTTVELRNLTATTGTLKLPATLVYDYPTTALADPELLAEELCA 6385
Qy 126 -----
Db 6386 LPESGAPAFVGRAADDPIVIGMNCRFPGGIRSPEDLWQLSSGEDAISGFPADRGMWL 6445
Qy 126 -----
Db 6446 DALARGASATLEGGLDGVGLFDAPFFGISPREALAMPQORLLLETSWEPERAGIDAT 6505
Qy 126 -----
Db 6506 TLRGSRGTVEVGTNGQDYATLLRQGTGTTDIRGHVATGNTASVMSGRLSYAFGLEGPVAT 6565
Qy 126 -----GACSP---
| | |
```

```
Db 6566 VDTACSSALVALHMAAGALRSGECTLALAGGVSVMSSPDSFTFTVQGLAPDGRCKPPA 6625
Qy 131 -----
Db 6626 DAADGTSWSEGVGLLERLSDARRNGHDVWGIVRGTAVNQDGASNGLTAPSGRAOORAI 6685
Qy 131 -----
Db 6686 RQALADADLAPADVVEAHGTGTTGLDPIEAHALIAAYQDRPTDRPLLLGAVKSNLGH 6745
Qy 131 -----
Db 6746 TORAAVAGVTIKVLLAMRHGVLPKLHTDPSAHVDMTACTVSLTTEQQEWPETGHARRA 6805
Qy 131 -----
Db 6806 GVSAGVSGTNAHVITEQAPDAESVAPEPTTVPGLVPMFVSGKSEAALEAGIERVSALT 6865
Qy 131 -----
Db 6866 GRSPVDVGHSLTTGTRAAFTHRAVLADGAGVRELARGVAHEADGKLAVLFSGOGAQRVGM 6925
Qy 131 -----
Db 6926 GRELYVRFPVFAEALDGLVLAHFDAGRDVMFGDAEGLDETGTPTOPALFAIEVALFRLAES 6985
Qy 131 -----
Db 6986 LGVRPDPFGSHSGEIAAAHVAGVFSLADACALVAABARLMOALPAGGAMVAVQATAEEA 7045
Qy 131 -----
Db 7046 AARLVDGVSIAAVNGPEAVVIAGEQREVLRIAKEFTGLGRKTRRLPVSHAFSPLMPEML 7105
Qy 131 -----
Db 7106 EDFERRVAEGLSYEAPRIALYNSLTGELATEDLRSABEYWRHVHRETIVRFADGVSAALAE 7165
Qy 131 -----
Db 7166 ASVFEELGPDGVLTAQAHTLOGTATAVPALRKDWSEETALLTALAQLHVVGVDWMSGI 7225
Qy 131 -----
Db 7226 FAGTGARRVDLPTYFQHQWYWDGAPAAEATTGAQDPVDAEFMDAVERADLSLSDSLE 7285
Qy 131 -----TWSCLI-----
Db 7286 LDDAALSALVPALSAMRRKRVSTVDGWRVYHVTFKPLTGTPOTLACTLWALVPTGNAD 7345
Qy 137 -----
Db 7346 DEWTTAVLAALGDAAVIAEVDPPADRGAFAGLGLAERGTEFTGVVSSQSRTSPSPSD 7405
Qy 137 -----
Db 7406 RGPTRRRWALAEAGIDAPLWCVTRDAVPVDRSESEVSPAQAAVWGLGRVAALDHPDRMG 7465
Qy 137 -----TETD-----
Db 7466 GLVDVPOEHRRTAERLRAVLSGAGDEDOVALRPSGAFGRRLVRATDDTTPATPHQAGTV 7525
Qy 141 -----
Db 7526 LVVGPASGTGHCARWLAGOGAQLVLADRTAPDTAQLIGELDALGVPLTVLDCAPADGE 7585
Qy 141 -----
Db 7586 ALLSALAALPEPLTAVIYADGPDADAYGTATTADPGASLAALYADVESLAABAAATGA 7645
Qy 141 -----
Db 7646 RPLGAFVLFSSIAGTWAGGRGTAAAGAOQLDALAGMLRARGLPATAVSWGWDATSGS 7705
| | |
```

QY	141	-----	140
Db	7706	LAHLRANGLPPMPDARALTALARSIGTDTSPVTADVVRWDEFAPAFTRRRHSALFAELP	7765
QY	141	-----	140
Db	7766	EAREALAESGSGAPGTADAASGURAEALARPDAERTEHLLALVRDVAAVLGFADADAVP	7825
QY	141	-----	140
Db	7826	SGQAFDLGLFDSLTAVDLRNLQNAVATGLALPATLVFDYPTADALAGYLRAELLGEGKDA	7885
QY	141	-----	144
Db	7886	EELPAVVRTKADTAEDPPVIVGMSCRYPGIRSPEDLMQVMGEADAIGFFPTDRGWDL	7945
QY	145	-----	144
Db	7946	DRLLHGRDRGRKTVTGGGFLYDVADFDGFFGIAPREANVMDPQORILLLEATHEALER	8005
QY	145	-----	144
Db	8006	TGIDPARLRGDTGVFIGGSGDYRPEAGQLGHAQTAQSASLLSGRVAVHFGLEGPSVSV	8065
QY	145	-----	148
Db	8066	DTACSSSLVALHAAQALNRNGECSIALTGCVTV	8098
RESULT 6			
Q09164			
ID	Q09164	PRELIMINARY;	PRT; 15281 AA.
AC	Q09164;		
DT	01-NOV-1996	(T-EMBLrel. 01, Created)	
DT	01-NOV-1996	(T-EMBLrel. 01, Last sequence update)	
DT	01-MAR-2002	(T-EMBLrel. 20, Last annotation update)	
DE	Cyclosporin synthetase (CYSYN) (EC 6.4.1.1).		
GN	SIMA.		
OS	Tolypocladium inflatum.		
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;		
OC	Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae;		
OC	Tolypocladium.		
OX	NCBI_TaxID=29910;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
RC	STRAIN=ATCC 34921;		
RX	MEDLINE=95094306; PubMed=8001164;		
RA	Weber G., Schoergerdorfer K., Schneider-Scherzer E., Leitner E.;		
RT	"The peptide synthetase catalyzing cyclosporine production in		
RT	tolypocladium niveum is encoded by a giant 45.8-kilobase open reading		
RL	frame."		
RL	Curr. Genet. 26:120-125(1994).		
CC	-1- FUNCTION: THE CONSTITUENT AMINO ACIDS OF CYCLOSPORINS ARE		
CC	ACTIVATED AS AMIONACYL-ADENYLATES WITH PEPTIDE BONDS FORMED		
CC	THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER INTERMEDIATES.		
CC	ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.		
CC	-1- COFACTOR: CONTAINS 11 COVALENTLY BOUND PHOSPHOPANTETHEINES.		
CC	-1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF CYCLOSPORINS.		
CC	-1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT		
CC	COVALENT BINDING OF AMP TO THEIR SUBSTRATE.		
DR	EMBL; Z283383; CAAB2227.1; --		
DR	HSSP; P14687; 1AMU.		
DR	InterPro; IPR000873; AMP-bind.		
DR	InterPro; IPR000267; Asp/Glutamase.		
DR	InterPro; IPR001242; Condensatn.		
DR	InterPro; IPR003880; Pantne attach.		
DR	InterPro; IPR000051; SAM_bind.		
DR	Pfam; PF00501; AMP-binding; 11.		
DR	Pfam; PF00668; Condensation; 13.		
DR	Pfam; PF00550; pp-binding; 11.		
DR	PRINTS; PR00154; AMPBINDING.		
DR	PRINTS; PR00139; ASGNLASE.		

DR	PROSITE; PS50075; ACP DOMAIN; 11.		
DR	PROSITE; PS00455; AMP BINDING; 10.		
DR	PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN 9.		
KW	Ligase; Antibiotic biosynthesis; Phosphopantetheine; Repeat;		
KW	Multifunctional enzyme.		
FT	DOMAIN 15179 15219	13 X 3 AA APPROXIMATE REPEATS.	
FT	REPEAT 3 1086	DOMAIN 1.	
FT	REPEAT 1087 2585	DOMAIN 2.	
FT	REPEAT 2586 4072	DOMAIN 3.	
FT	REPEAT 4073 5564	DOMAIN 4.	
FT	REPEAT 5565 7061	DOMAIN 5.	
FT	REPEAT 7062 8121	DOMAIN 6.	
FT	REPEAT 8122 9616	DOMAIN 7.	
FT	REPEAT 9617 11113	DOMAIN 8.	
FT	REPEAT 11114 12185	DOMAIN 9.	
FT	REPEAT 12186 13681	DOMAIN 10.	
FT	REPEAT 13682 14767	DOMAIN 11 (ALA-ACTIVATING).	
FT	BINDING 1060 1060	PHOSPHOPANTETHEINE (POTENTIAL).	
FT	BINDING 2558 2558	PHOSPHOPANTETHEINE (POTENTIAL).	
FT	BINDING 4045 4045	PHOSPHOPANTETHEINE (POTENTIAL).	
FT	BINDING 5537 5537	PHOSPHOPANTETHEINE (POTENTIAL).	
FT	BINDING 7034 7034	PHOSPHOPANTETHEINE (POTENTIAL).	
FT	BINDING 8094 8094	PHOSPHOPANTETHEINE (POTENTIAL).	
FT	BINDING 9589 9589	PHOSPHOPANTETHEINE (POTENTIAL).	
FT	BINDING 11086 11086	PHOSPHOPANTETHEINE (POTENTIAL).	
FT	BINDING 12158 12158	PHOSPHOPANTETHEINE (POTENTIAL).	
FT	BINDING 13654 13654	PHOSPHOPANTETHEINE (POTENTIAL).	
FT	BINDING 14729 14729	PHOSPHOPANTETHEINE (POTENTIAL).	
SQ	SEQUENCE 15281 AA; 1689051 MW; E26DA7AA35324C05 CRC64;		
Query Match 34.2%; Score 264; DB 3; Length 15281;			
Best Local Similarity 0.8%; Pred. No. 4.8;			
Matches 102; Conservative 24; Mismatches 21; Indels 12475; Gaps 32;			
QY	1	AAGGIL-----	6
Db	2098	AAGHVLEIGTGTGMVLFNLQGLKSYIGLEPSQSAVFNKAAOTFFGLEKGAQVHGT	2157
QY	7	-----	6
Db	2158	AMDTGRLSALSPDLIVNSVAQYFPRSREYLAEVLEALVRIPOVRRIFFGDMRTYATHKDF	2217
QY	7	-----	11
Db	2218	LVARAVHTNGSKVTRSKVQOEVARLEELLEDVDPAPFTSLKESLSEIEHVEILPKNM	2277
QY	12	-----	11
Db	2278	KYNELSSYRYGAVLHIRNHNNQNSRSHKNAESWIDPASSQMDROGLARLLKENKDAE	2337
QY	12	-----	11
Db	2338	STAVFNIPYSKTIIVERHIAKSLADDDHGDGDDTHSSIDGVAWISAAAREKASQCFSLDVHDLV	2397
QY	12	-----	11
Db	2398	QLAEDAGFRVEYSWARQSRONGALDVFFHHFQPTENESRALVDFPTDYKGOQARSLTNRP	2457
QY	12	-----	11
Db	2458	LORVESRRIEAOVREQLQVLLPAYMIPARIVLQNMPLNTSGKVDKRLTLRAKVTAART	2517
QY	12	-----	11
Db	2518	PSSELVAPRDSIEAIIKCFKDVGLGVEVGITDNFFNVGGHSLATKLAARLSQLNAQIA	2577
QY	12	-----	11
Db	2578	VKIDFDRPVIADLAATIQQDTEHNPILTSTYGPVQSFQAGRLWFLDQLNVGATWYLM	2637
QY	12	-----	11
Db	2638	PFAVRLRGPLVWSALAAALLAERHETLRTTTFIEQEGIMQVHPFAPKELRVIDVSGE	2697

QY	12	-----	11	QY	19	-----	18
Db	2698	ESTIQKILEKEQTPPNLASEPGRLLALLKTGEDEHILSTMHHAISDGHWSVDIFQOEI	2757	Db	3778	SYRYAAVLHVNDLAKPAHKVSPGAWVDFAAKTMDRDALIRLLRGTKISDHIAIANIPNSK	3837
QY	12	-----	11	QY	19	-----	18
Db	2758	QOFYSAILRGHDPLAQIAPLSIQYRDFATWQROIQFQVAEHRRLAYMTKQLADNKPABLL	2817	Db	3838	TIVERTICESVYDLGGDAKSDNRVSWLSAARSNAVKVASLSAIDLVDIAOAGFRVEIS	3897
QY	12	-----	11	QY	19	-----	18
Db	2818	TDFKRPMLSGRAGEIPVYVBDGLIYEKLODFCRIRQVTAFTVLLAAFPAAHYRMTGTEDA	2877	Db	3898	CARQWSONGALDAVPHHLGSPSPOSSHVLIDFLTDHQGRPEEALTNHPLHRAQSRRRVERQI	3957
QY	12	-----	11	QY	19	-----	18
Db	2878	TIGTPIANRRPELEGLIGFFVTCMRITVDVEDSFETLVHQVRETTLAHAHANQDVPFE	2937	Db	3958	RERLOTLLPAYMIPAQIMVLDKPLNANGKVRDKOUTQRAQTVPKAKQVSAPVAPRTEIE	4017
QY	12	-----	11	QY	19	-----	18
Db	2938	QIVSNILPGSSDTSRNPVLQMFALHSQNLGKVRLEGIEEBEIIISIAETTRFDIEFHLYQ	2997	Db	4018	RVLCOFSFSDVLGVDIGIMENFFDLGGHSLMATKLAARISRRLETHSVSKEIFDHPRVCDL	4077
QY	12	-----	11	QY	19	-----	18
Db	2998	EAERLNGSIVVAADLFVPETIQSVITIFQILQKGLGEPDMPVSMALDGGLESRLSTGL	3057	Db	4078	VLIVQOGSAPHDPVSTKYTGVPQSPAQGRWFLDQLNFGATWYLMPLAVRLRGAMNVH	4137
QY	12	-----	18	QY	19	-----	18
Db	3058	LHQOQTDYPCDASVVOIFKQVAVNPDIIVRDESTRLSYADLDRKSDQVACWLSRRGIA	3117	Db	4138	ALTAALLALERHRELLRTTFYEQNGVMQKVPVVTETURIIDLSNGDGDYLLTKKEQT	4197
QY	19	-----	18	QY	19	-----	18
Db	3118	PETFAVAILAPRSCETIVAILGVKANKANLAYPLDVNPVAPASLEAILSEVSGSMLVLGAET	3177	Db	4198	APHLETEPGWRVALLRGLPGDYILSVMMHIIISDGSVDVLFQELGQFPYSTAVKGHDPL	4257
QY	19	-----	18	QY	19	-----	18
Db	3178	PIEGMAEATIRITEILADAKTDDINGLAASQPTAASLAYVFTSGSTGRPKGVMEHR	3237	Db	4258	SOTTPLPIHYRDPALWQKPTQSEHERQLQYWVEQLVDSAPAEILLTDLPRPSILSQAG	4317
QY	19	-----	18	QY	19	-----	18
Db	3238	GIVRLTKQNTSKLPESFHMAHISNLAFDASVWEVFTLLNGGTLVCIDIFTLLLESTAL	3297	Db	4318	EMSVTIEGALYKNLEEFRCVRHVTSFVLLAALRAAHYRLTGSSEDATIGTPIANRRPEL	4377
QY	19	-----	18	QY	19	-----	18
Db	3298	EKVFFDQRVNVALLPPALLKQCLDNSPALVKTLSVLYIGDRLDASDAKARGLVQTAQF	3357	Db	4378	EOIIGFFVNTQCIRITVNEDETFESLVQOVRSTATAAFAHQDVPFEKIVSTLLPGSRDAS	4437
QY	19	-----	18	QY	19	-----	18
Db	3358	NAYGPTNTVMSTIYPIAEDPFIINGVPIGHAVSNSCAFVMDQNOQITPPGANGELIVTGD	3417	Db	4438	RNPLVQLMFAVHSQKNLGELKENAHSEVVPTEITTRFDLEFHLFOODDKLEGSILYSTD	4497
QY	19	-----	18	QY	19	-----	18
Db	3418	GLARGYTTSSLNTRFINVDIDGQVRYRTGDRVRYRPKDLQIEFFGGRIDHGVKIRGHR	3477	Db	4498	LFEAVSVQSLLSVFQEILRRGLNGPDVPISTLPLQDGIVDLQROGLLDVQKTEYPRDSSV	4557
QY	19	-----	18	QY	19	-----	18
Db	3478	IEPAEVEYALLSHDLVDAAVVTHSQENQDLEMVGFVAARVADVREDESSNOVQEWQTHF	3537	Db	4558	VDVPHEQVSNPDSIALIHGSEKLSYAQLDRESDRVARWLHRHSFSDTLIAVLAPRSC	4617
QY	19	-----	18	QY	19	-----	18
Db	3538	DSIAYADITTIQQSLGRDFMSWTSNMYDGLIKKSQOEMLDDTMRSLDSQPPGHVLEV	3597	Db	4618	TIIFLGLKANLAYPLDVKAPAARIDAIVSSLPGKNKILLGANVTPPKLOEAAIDFVP	4677
QY	19	-----	18	QY	19	-----	18
Db	3598	GTGTGMVFNLCREGGLQSVGLEPSPSATAFVNKAAKSFPGLEDRIRVEVGTATDIORL	3657	Db	4678	IRDTFTTLTDGTLDQDGTIERPSAQSLAYAMFTSGSTGRPKGVMMVQHRNIVRLVNSNVV	4737
QY	19	-----	18	QY	19	-----	23
Db	3658	GDDLHAGLVVNSVAQYFSPDYLAQLVRDLTKVPGVERIFFGDMRSHAINRDFLVARAV	3717	Db	4738	AKOPAAARIAHISNLAFDASSWEIYAPLLNGAIVCADYFTTIDPOALQETFOEHEIRGA	4797
QY	19	-----	18	QY	24	-----	23
Db	3718	HALGDKATKAEIOREVVRMESEDELLVDPAPFTSLTTOVENIKHVIEILLPKMRATNELS	3777	Db	4798	MLPPSLLKQCLVQAPDMIISRLDILFAAGDRFSSVDALQAOQLVGVGVFNAYGPTENTILS	4857
				QY	24	-----	23

Db	4858	TIYNVAENDSFVNGVPIGSAVNSGAYIMDKVQQLVPAGVMGELVVTGGDLARGYMDPKL	4917
Qy	24	-----	23
Db	4918	DADRFIQLTVNGSEQVRAVRTGDRVYRPFKQIEFFGFRMDQOIKIRGHRIBPAEVEQAF	4977
Qy	24	-----	23
Db	4978	LNDGFVEDVAIVIRTPENOPEMFAVFTAKGDSNAREEBATTQIEGWEAHFEGGAYANIE	5037
Qy	24	-----EDT-----	26
Db	5038	EIESEALGYDFMGWTSWYDGTEDKDEMREWLNDTMRSLDDGKPAGRVLEVGTGCMIMF	5097
Qy	27	-----BRYV-----	30
Db	5098	NLGRSGLERYIGLEPAPSAAEFVNNAKSFPGLAGRAEVHVGTAAVGTLOGLTSDMAV	5157
Qy	31	-----	30
Db	5158	INSAQYFPTPEYLAETIKSLVQVPGMKRIYLGDMRSMWNRDFAAARAAYSLADNASKD	5217
Qy	31	-----	30
Db	5218	RVRQKMELEKEEELLVDPAFTALASQLODRIQHVEILPKMKATNELSSRYAAVLH	5277
Qy	31	-----	30
Db	5278	ISDEPLPIYKIDPEAMINPEGSRLTREALAQVLKENENAEVAISNIPYSKTVVERHIVR	5337
Qy	31	-----	30
Db	5338	SLDQEDANAPESMGSDMISAVTRAQOCHTSLASDLDIAEDAGFRVEVSWARHOSH	5397
Qy	31	-----LTN-----	33
Db	5398	GALDAVPHHLKPATEDSRVLKPTDHOGRPLKSLTNQPLLPQASRRAEILLIREGLQTL	5457
Qy	34	-----	33
Db	5458	PPYMIPOITLIDRMLNANGKVDRELARRAKITOKSKPEDIVPPRNSVEATVCKGFT	5517
Qy	34	-----	33
Db	5518	DVLGVEVGITDNFNLGGHSLMATKLAARLGRQINTRISVRDVPDQVPVADLAIVQRNS	5577
Qy	34	-----LNIGA-----	38
Db	5578	APHEPIKPADYTGVPQSPAGRLWFLDQNVGATWYLMPLGIRLHGLSRVDALATAISA	5637
Qy	39	-----	38
Db	5638	LEQRHEPLRTTFHEEDGVGVQVQDHRPKDLRIIDLSTQPKDAYLAVLKHEQTLFDLAT	5697
Qy	39	-----	38
Db	5698	EPGMRVALIRLGBEEHILSIWMHHIISDGWSVEVLFDEMHRFPYSALRQODPMEQILPLP	5757
Qy	39	-----	38
Db	5758	IOYRDAFAWQKTEQVAEHQOLDYWEHLADSTPAELLTDLPRPSILSGRANELPLTIE	5817
Qy	39	-----	38
Db	5818	GRLHDKLRAFRCRVHOATPFVILLAAAHYRLTGAEDATLGTPIANRNRELENMIGFF	5877
Qy	39	-----	38
Db	5878	VNTQCMRIAIBENDNFESLVRVRSTATSAFANQDVPFESIVSLLPGSRDASRNPLVQV	5937
Qy	39	-----	38
Db	5938	ILAVHSQDGLKLTLEGLRDEAVDSAISTRDFVEFHLEHADRLSGSVLYAKELFKLRTI	5997
Qy	39	-----	38
Db	5998	ESVSVFLETLRRALDQPLTPLAVPLTDGVEIASKGLLDVPRDTPDYPRDANIVEVFOQH	6057
Qy	39	-----	38
Db	6058	VRATPDAIVKDATSILTYAQLDQSDRLAIWLSRRHMPETLVGLVAPRSCETIIAMPG	6117
Qy	39	-----	38
Db	6118	IMKANLAYLPDINSIPAARLSILSAVDGNKLVLLGSGVTAPEQENPEVEAVGIOEILAG	6177
Qy	39	-----	38
Db	6178	TGLDKTQGSNARPSATSLAYVIFTSGTGPKGVMEHRSVTRLAKPSNVISKLPQGARV	6237
Qy	39	-----	38
Db	6238	AHLANIAFDASIWEIATTLNGATLVCLDYHTVLDCTRLKEVFERESITVVTLMPALLQK	6297
Qy	39	-----	38
Db	6298	CVAEIPETLAHLDLLYTGDRVGGHDAMRARSILVKGHFGSGYGTENTVISTIIYEVDAD	6357
Qy	39	-----	38
Db	6358	MFVNGVPICKTVSNGAYVMDRNOQLVPSGVVGGELVVTGDLARGYTDPSLNKNRFIVIT	6417
Qy	39	-----	38
Db	6418	VNGESIRAYRTGDRVYRPHDQIEFFGRMDQOVKIRGHRIEPEGEVESALLSHNSVQDAA	6477
Qy	39	-----	38
Db	6478	VVICAPADODSGAEMVAFVAARNTEDTQEEAVDQVQGWETHFETAAYSEVKDIROSE	6537
Qy	39	-----	38
Db	6538	VGNDPMGWTSMYDGEIDKTDHMLNDTMRMILDAREPHGVLEIGTGTGMVFNLAACP	6597
Qy	39	-----	38
Db	6598	GLQGYVGFEPKSAQFVNDAAQSPALKDGRSIVHVGTTATDINKAGPIOPRLVVINSVA	6657
Qy	39	-----	38
Db	6658	QYFPTPEYLFVVVEALVQIPSVVERIVFGDMRTNAINRDFVASRALHTLGEKANKRLVRQM	6717
Qy	39	-----ELLRDPS-----	45
Db	6718	IYELEANEHEELTDPAPFTSLRGLGEKIKHVEILLPKMTKATNELSKRYAAVLHVGRSR	6777
Qy	46	-----	45
Db	6778	EQSTIHQVSPNAWIDPAADGLDROTLLINLLKEHKOAGTVAIGNIPYSKTIVERFVNKSLS	6837
Qy	46	-----	45
Db	6838	EDDMEEQNSLDGSAWAAVRAAQAQSCPSLDAMDVKEIAQEAQYQVEVSWARQWSONGAL	6897
Qy	46	-----LCAQ-----	49
Db	6898	DAIFHHFEPKEGARTLIEFPTDYEGRVNVTLTNRPLNSIOSRRLLGTQIREKLOTLPPY	6957
Qy	50	-----	49
Db	6958	MIPSRIMVLDQMPVNNNGKIDRKELVRRRAIVAPKPSAATRVAPRNEIEAILRDEFEDVL	7017
Qy	50	-----	49
Db	7018	GTEVSVLDNPFDDLGGHSLMATKLAARVSRRLDAHISIKDVFQDQVLADLAASIQRESAPH	7077

Qy	50	-----	49	Qy	74	-----	73
Db	7078	EPQPQYTPAQSFQAGRLWFLDQNLGATWYLMPLAIRGOLRVAALSAALFALER	7137	Db	8158	LWFLEQLNLGASWYLMFPFAIRMRGPIQTKALAVNALVHRHEALRTTFEDHGVQV1	8217
Qy	50	-----	49	Qy	74	-----	73
Db	7138	RHETLRTTFERSDGVQVQVGEARNSDLRVHDSVSTGDDGGEYLEVLRRQTVPFDLSSPEP	7197	Db	8218	QPKSSODLR1IDLSDAVDDTAYLAALKREOTTAFDLTSEPGWRVSLRLRGDDDDYILSVM	8277
Qy	50	FRVHLVK-----	56	Qy	74	-----	73
Db	7198	WRVCLVKTEEDHVLISVMHHIYDGSVDILRGELGQFYSAALRGQDPLHANPLPIQY	7257	Db	8278	HHIISDGTVDVLRQELGQFYSAAIRGOEPLSOAKSLPIOVDFAVWQROENIQEAKQ	8337
Qy	57	-----	56	Qy	74	-----	73
Db	7258	RFAAWOREAKQVEEHQOLGW\$KOLVDSTPAELLTDLPRPSILSGRAGSDVDTIEGSV	7317	Db	8338	LKYWSQOLADSTPCFELTDLPRPSILSGEADAVPMVIGDTVYQLLTDPCRTHOVTSFVL	8397
Qy	57	-----	61	Qy	74	-----	73
Db	7318	YGALQSFCTRSTVTFVLLTVFRIAHFRLTAVDATTGTPIANRNRPELETLVGCFVNT	7377	Db	8398	LAAFRTAHYRLTGLTATVGTPTIANRNRPELEGLIGFVNTQCMRAISETETESLVQQ	8457
Qy	62	-----	61	Qy	74	-----	73
Db	7378	OCMRISADDDNFEGLVROVRNVATAAVANQDVPFERIVSALVPGSRNTRSRLVQLMFA	7437	Db	8458	VLTTTEAFANQDVPFEQIVSTLLPGSRDTSRNPLVQVMFALOSQODLGRIGLEMTDEA	8517
Qy	62	-----	61	Qy	74	-----	73
Db	7438	VOSVEDYQOVRLEGLESVMPEGASTRDMFHLVPGDKLTGSLVYSSDLPQEQTIONF	7497	Db	8518	LETPLSTRDLVHLFQEVGKLSGLLYSTDLFEVETIRGIVDVFLERLRRGLEBQKORL	8577
Qy	62	-----	61	Qy	74	-----	73
Db	7498	VDIFOECLRSVLDQPLTPISVLFPFSNAISNLESLDLLEMTSDYPRDRTVVDLFRQAAI	7557	Db	8578	MAMPITDGITKLROGILLTAKPAYPRESSVIDLPROQVAAAPDAIAVWDSSSLTLYADL	8637
Qy	62	-----	61	Qy	74	-----	73
Db	7558	CPDSIAVKDSSSOLTVAOLDEQSDRVAAMLHERHMPAESLVGLSPRSCETIIAYFGIMK	7617	Db	8638	DCQSNKLAHMLCQRNMAPETLVAVFAPRSCLTIIVAPLGVLKANLAYLPDYNAPARIEA	8697
Qy	62	-----	61	Qy	74	-----	73
Db	7618	ANLAYPLDVYAPDARLAAILDVGERLLLLGAGVQPQIQIPLRSTAYIAEALSHATT	7677	Db	8698	ILSAVPGHKLVLQAHGPELGTWADTELVOIDBALASSSGDHEQIHASGPTATSUYV	8757
Qy	62	-----	65	Qy	74	-----	73
Db	7678	VDVTSIQPSATSALAVIFTSSTGPKGVMIEHRGIVRLVDTNVNVPFSGSALPVSH	7737	Db	8818	GCTLVCLDYLTLDSKILYNNVFEAQVNAAMFTPVLLKQCLGNMPAIIISRLSVLFNVGDR	8877
Qy	66	-----	65	Qy	74	-----	73
Db	7738	FSNLWDAATWEIYTAVLNGGTVCICDRDTMLDIAALNSTFRKENVRAAFFTPAFLKQCL	7797	Db	8938	GNQQLVSPGVNGEIVVTGDLARGYTDSALDEDRFVHTIDGEEENIKAYRTCDRVRYRPK	8997
Qy	66	-----	73	Qy	74	-----	73
Db	7798	AETPELVANLEILHTAGDRLDPGDANLAGKTAKGGIFNVLGHTEHTAYSTFFYPVVGEEFT	7857	Db	8998	DPEIEFFGRMDQOVKIRCHRIEPAEVEHALLGDLVHDAAVVLRKPAQEPPIAFITSO	9057
Qy	74	-----	73	Qy	74	-----	73
Db	7858	VNGVPVGRGISNSHAYIIDRHQKLVPAVGMGELLTGDGVARGYTDNALNKDRFVVYDIN	7917	Db	9058	EDETIEOHESNKQVQGWGEHFDVSRYADIKOLDTSTFGHDFLGTWTSMDGVVDIPVNEKE	9117
Qy	74	-----	73	Qy	74	-----	78
Db	7918	GKSTWYRTGDKARYRPRDQGLEFFGRMDQMKIRGVRIEPEGEVELTLLDHKSVALAATVV	7977	Db	9118	WLDETTASLLDNRRPPCHILEIGAGTGMILSNLKGKVDGLQKYVGLDPAPSAI FVNEAVKS	9177
Qy	74	-----	73	Qy	79	-----	78
Db	8038	DRKDLALRAQTVQKRRTAARVPRDEVEAVLCEEYSNLLVEVEVGITDGFDLGCHSLLA	8097	Db	9178	LPSLACKARVLVGTALDIGSLDKNEIQPELVVINSVAQVFTSEYLLIKVWKAWEVPSVK	9237
Qy	74	-----	73	Qy	79	-----	78
Db	8098	TKLAARLSQLNTRVSKVDFOPIILADLADIIRRGSHRHRDIPATPYTGPVEQSFQAGR	8157				

Db	9238	RVFGDIRSQALNPDFLAARAVRALGDNASKEQIREKIAELESEEBELLVDPAFFVSURS	9297
Qy	79	-----	78
Db	9298	QLPNIKHVEVLPKLMKATNELSSRYAAVLHISHNEEBQLLIQIDPTAWVDFAATOKDS	9357
Qy	79	-----SVC 81 :	
Db	9358	QGLNLLQQGRDDVMIAGNIPYKTIIVERHIMNSLDQHVNSLDGTSWISDARSAAAIC	9417
Qy	82	-----	81
Db	9418	TSPDAPALTQAKKEGFRVELSWARQSONGALDAVPHRLATDANCERSRVLVHPTDHO	9477
Qy	82	-----	81
Db	9478	GRQLRTLNRPLORAQSRRIESQVFEALQATALPAYMIPSRRIIVLPOMPTNANGKVDKQOL	9537
Qy	82	-----	81
Db	9538	ARRAQVAVAKKAVSAPVRNDTEIVLCEEYADILGTEVGITDNFDMGGHSLMATKLA	9597
Qy	82	-----	81
Db	9598	RLSRRLDTRVTVKEVDFDKPVLADLAASIEQGSTPHLPASSVYSGPVEQSYAQGLMFLD	9657
Qy	82	-----	81
Db	9658	QFNLNATWYHMSLAMRLLGPLNMDALDVALRALQORHETLATTFEAQKDIGVQVVHEAGM	9717
Qy	82	-----	81
Db	9718	KRLKVLDSLCKNEXEHMAVLEQMRPPTLASEPGWKHGLARLGPTIILSLVMHMFSD	9777
Qy	82	GWS----- 	84
Db	9778	GWSVDILRQELGQFYSAAALRGDRPLSQVKPLPIQYRDFAAWQKEAAQVAEHERQLAYWEN	9837
Qy	85	-----	84
Db	9838	QLADSTPCELLTDFRPOFLSGKAGVIPVTIEGPPVYEKLLKFSKEROVTLFSLVLLTAFRA	9897
Qy	85	-----	84
Db	9898	THFRLTGAEDATIGTPIANRNPDEHIIGFFVNTQCMRLLLDTGSTFESLVQHVRSVAT	9957
Qy	85	-----	84
Db	9958	DAYSNODIPFERIVSALLPGSRDASRPLIQLMFALHSQPOLGNITLEGHERLPTSVA	10017
Qy	85	-----QTIN----- :	88
Db	10018	TRFDMEFHFOEPNKLSCSILLFADELFOPETINSVVTFQEILRGLDQPOQVISTMPLT	10077
Qy	89	-----	88
Db	10078	DGLIDLEKLGLEIESSNFPDYSVVDVFRQOVAANPNAPAVVDSETSMTSLDQKSEQ	10137
Qy	89	-----PE----- 	90
Db	10138	IAAMLHAQGLRPESLICVMAPRSFETIVSLFGILKAGAYPLPDVNSPAARIQPIELSE	10197
Qy	91	-----	90
Db	10198	GKRLVLLSGGDMPOSMDVETARIQDILTNTKVERSDPMSRPSATSLAYVFTSGSTG	10257
Qy	91	-----	90
Db	10258	RPKGVMIHRNTRLRLVKOSNVTSQLPDQLRMAHISNLAFDASIWEITAILNGGALICID	10317
Qy	91	-----	90

Db	10318	YFTLLDSQALRTTTFEKARVNA TLFAPALLKCELNHAPTLPEDULKVLYIGGDRLDATDAK	10377
Qy	91	-----90	
Db	10378	IQALVKGTVYNA GPTENTVMSTIYRLTDGESYANGVPIGNVSSGAYIMDKQRLVPP	10437
Qy	91	-----90	
Db	10438	GVMGELVVGDLARGYTNSTLNADRFVDIVINDQKARAVRTGDRTRYRPGKGSIEFFGR	10497
Qy	91	-----90	
Db	10498	MDQVQKIRGRHVEPAEVEQAMLGKAIHDAVVVQVAVDQGTEMIGFVSMASDRFSGEE	10557
Qy	91	-----DDT---93	
Db	10558	EITNOVQEWEDHFESTAYAGIEADQATLGRDFTSWTSMYNGNLI DKAEMEELDDTMO S	10617
Qy	94	-----93	
Db	10618	LLDKEDARPCAEIGTGTGMVLNLPKNDGLESYVGIEPSSAALFVDKAAODFPFGLOKT	10677
Qy	94	-----93	
Db	10678	QILVGTAEEDIKLVKDFHPDVPVVIINSVAQYFPSRSYLVQIASELIHMTSVKTIFFGDMRSW	10737
Qy	94	-----DP-----95	
Db	10738	ATNRDLVSALYTLGDKATKDQIQOEVARLEENEDELLVDPAFFTSLTSQPGVKHVE	10797
Qy	96	-----95	
Db	10798	ILPKMRMTSNELSSYRYAAVLHI CRDGEGRNRYRRVHVSVEENAWIDPASSGMDRHALVQ	10857
Qy	96	-----95	
Db	10858	MLDERRDAKTVAIGNIPHNSNTINERHFTSLDTEGEGIAQDSLDCSAWQSAWKAMAARCP	10917
Qy	96	-----95	
Db	10918	CLSVTELVEIGGAAGFRVEVSWARQSOHGALDVVFHHLDDRVGRVLINFPDTDFERLPP	10977
Qy	96	-----95	
Db	10978	STGLTSRPLQIONRRRPFESIREQLTLLPPYMPVSRIVVLERMLPNANSKVDRKELARK	11037
Qy	96	-----GH-----97	
Db	11038	ARTLOTIKPSATRVAPRNDIEAVLCDEFOAVLGVTVGVMDNPFELGGHSLMATKLAARLS	11097
Qy	98	-----ADLV-----101	
Db	11098	RRLDTRVSKDIFNQPILODLADVVGTSAPHEAIPSTPYSPVQEQSFSGRLWFLDQLN	11157
Qy	102	-----101	
Db	11158	LNASWYHMPLASRLRGLRIEALQSALATIEARHESLRTTFEODGVPVQIVRAARNKOL	11217
Qy	102	-----101	
Db	11218	RIIDVSGTEDAYLAALKOEQDAAFDLTAEPGRVRAALLRGLGDDHVLSTVMHHIISDGHSV	11277
Qy	102	-----LY-----103	
Db	11278	DILROELGOLYNASSQAPLPIQYRDAIMQKDSQIAEHOKQLNWKROLVNSKPAEL	11337
Qy	104	-----103	
Db	11338	LADFTRPKALSGDADVPIEIDDDQVYQNLRSFCRARHVTFSVALLAFAAAHYRLTGAED	11397
Qy	104	-----103	
Db	11398	ATIGSPIANRNPPELGLIGCFVNTQCLRIPVKSEDTFTDLVKQARETATEAODNQDVPF	11457

Qy 104 -----ITRFDLEL--- 111
Db 11458 ERIVSSMVASSRDTSRNPLVQVFAVHSQHDGNIRLEGVEGKPVSMASSTRFPAEMHLF 11517
Qy 112 ----- 111
Db 11518 EDQMLGNVVFVKDLFESETIRSVAVFOETLRRGLANPHANLATPLTDGLPSLRSLC 11577
Qy 112 ----PD----- 113
Db 11578 LOVNQDYPDASVIDVPREQVASIPKSIADASSOLTYTELDERSSQATWLRQVTV 11637
Qy 114 -----GNQOV----- 118
Db 11638 PEELVGLAPRSCETIIAFLGIIKANLAYLPDYNAPAGRIETILSLPCNRLILGSDT 11697
Qy 119 -----R 119
Db 11698 QAVKLHANSVRFRTRISDALVESGSPTEELSTRPTAQSLAYVMFTSGTGVPKGVMEHR 11757
Qy 120 GVT----- 122
Db 11758 GITRLVKNVNAKOPAAAAIAHLSNIAPDASSWEIYAPLLNGTVVCIDYYTTIDIKAL 11817
Qy 123 ----- 122
Db 11818 EAVFKQHIRGAMLPPALLKQCLVSAPTMISSLEILFAAGDRSSQDAILARRAVSGVY 11877
Qy 123 ----- 122
Db 11878 NAYGPTENTVLSTIHNIGENAEFSGVPIGNVNSGAFVMDQONQQLVAGVIGELVWTG 11937
Qy 123 ----- 122
Db 11938 DGLARGYDTSKLRVDRFIYITLDGNVRAYRTGDRVHRPKQOIEFFQGMQQIKIRGH 11997
Qy 123 ----- 122
Db 11998 RIEPAVEQALARDPAISDAVITQUTDBEEPELVAFSLKGNANGTNGVSDQEKID 12057
Qy 123 ----- 122
Db 12058 GDEQHALLMENKIRHNLQALLPTYMIPRIIHVDQLPVNANGKIDRNEIAVRAQATPRTS 12117
Qy 123 ----- 122
Db 12118 SVSTYVAPRNDIETIICKEFADILSVRVGIDTNFDFLGGHSLIATKLAARLSRRLDTRVS 12177
Qy 123 ----- 122
Db 12178 VRDVFDPVVGOLAASIOQGSTPHEAIPALSHSGPVQOSFAQGRFLWFLDRFNLAAYIM 12237
Qy 123 ----- 122
Db 12238 PFGVRLGRPLVDALOTALALEERHELLRTTFFEQDQGVGMQIVHSPMRDICVVDISGA 12297
Qy 123 ----- 122
Db 12298 NEDLAKKEEQQAPFNLSTEVAVRVALFKAGENHHLSIVMHHIISDCGWSVDIFQOELAQ 12357
Qy 123 ----- 122
Db 12358 FYSVAVRGHDPLSQVKPLPIHYRDPFAVWORQKQAVHESQLQYWIEQADSTPAEILSD 12417
Qy 123 ----- 122
Db 12418 FNRPEVLSGEACTVPIVIEDEVYKLSLFCRNHQVTSFVLLAAPRAVHAHYRLTGAEDATI 12477
Qy 123 ----- 122
Db 12478 GTPIANRNRPELEDLIGFFVNTQCMRIALEEHDNPLSVVRRVRSSTAASAFENQDVPERL 12537

Qy 123 ----- 122
Db 12538 VSALLPGSRDASRNPLVQLMFVHVSORNLGKLEGESEPTPYTATTTRDFVEFHLFEOD 12597
Qy 123 ----- 122
Db 12598 KGLAGNVVFAADLFEAATIRSVVEVFHEILRRGLQDPDIAISTMPLVDGLAALNSRNLP 12657
Qy 123 ----- 122
Db 12658 VEDIEPDPFATEASVVDVFQTVVANPDALAVTDTSTKLTYAELDOOSDRAAWLSKQKLP 12717
Qy 123 ----- 122
Db 12718 ABSIVVVLAPRSSETIVACIGILKANLAYLPWDSNVPEARQAIISEIPGEKVFLLGAGV 12777
Qy 123 ----- 122
Db 12778 PIPDNKTADVVRMVFISDIVASKTDSYSGPTRPSASSLAYVIFTSGSTGRPKGVMEHRG 12837
Qy 123 ----- 122
Db 12838 VISLVKQNASRIPOSLRMAHVNLAFDASVMEIFTLLNGGTLFCISYFTVLDKALSAA 12897
Qy 123 ----- 122
Db 12898 FSDHRINITLLPALLKOCLADAPSVLSSLESLEYIGGDRLDGADATKVVDLVKGKAYNAY 12957
Qy 123 ----- 122
Db 12958 GPTENSVMTIYTIETIHFETANGVPITGSLGPKSKAYIMDDOQLVPAGVMGELVAGDGL 13017
Qy 123 ----- 122
Db 13018 ARGYTDPSLNTGRFIHITIDCKQVOAYRTGDRVRYRPRDYOIEFFGRLDQOIKIRGHRIE 13077
Qy 123 ----- 122
Db 13078 PAEVEQALLSDSSINDAVVSAQNKEGLEVMGYITTOAQSDVDEEASNKVQWEAHFDS 13137
Qy 123 ----- 122
Db 13138 TAYANIGIDRDALGDFLSWTSWYDGLIPREEMOEFLNDTMRSLLDNOPPGKVLBEIGT 13197
Qy 123 ----- 122
Db 13198 GTGMVLFNLKVEGLQSYAGLEPSRSVTAWNKAIEFFPSLAGSARVHVGTAEIDISSIDG 13257
Qy 123 ----- 122
Db 13258 LRSLDVVINSVAQVFPSPREYLAELTANLIRLPGVKRIFFGDMRTYATNKDPLVARAVHTL 13317
Qy 123 ----- 122
Db 13318 GSNASKAMVROQVAKLEDEBELLVDPAPFTSLSDQFPDEIKHVEILPKRMAATNELUSSY 13377
Qy 123 ----- 122
Db 13378 RYAAVHVGHCHOMPNGEDEDKQAVKDINPKAWVDFACTMDRQALLQLLQDRQRGDDVV 13437
Qy 123 ----- 122
Db 13438 AVSNIPYSKTIEMERHLSQSLDDEDDGTSVADGTAMISRTQSRACEPCALSVADLIEIGK 13497
Qy 123 ----- 122
Db 13498 IGFEVEASWARQHSORGGLDAVPHRPEPPRHSCHVMFRPTEHKGRSSSLTNRPLHLLO 13557
Qy 123 ----- 122
Db 13558 SRRLEAKVRERLQSLLPYMIPSRITLLDQMLPSTNGKYDRKKLARQARVIPSASASTLD 13617
Qy 123 -----OLGG----- 126

Db 13618 FVAPRTEIENVLCEEFTDLGVGVGIDTNFFELGSHSLATKLSARLSRRLDAGITVKQV 13677
QY 127 ----- 126
Db 13678 FDQVFLADLAASILQSSRRHSIPSPYEGVEQSFQAQRLWFLDQFNIDALWYLIPFAL 13737
QY 127 ----- 126
Db 13738 RMRGPLOVDALAAALVALEERHESLRITFEERDGVGIQVQVPLRTTKDIRIIDVSGMRDD 13797
QY 127 ----- 132
Db 13798 DAYLEPLQEQQTDFDLASEPFGWRVALLKLGKDDHILSIVMHHIISDGMSTEVLORELQ 13857
QY 133 ----- 132
Db 13858 FYLAASKGAPLQVAPLPIQYRDFAVWQREOVAESQRLDYWKQOLADSSPAELLAD 13917
QY 133 ----- 132
Db 13918 YTRPNVLSGEAGSVFVINDSVYKLSVFCRSQVTTFTTLAFAAAHYRMTGSDDATI 13977
QY 133 ----- 136
Db 13978 GTPIANRNPENLIGCFVNTQCMRITIGDDEFSLVQVQVSTTTATAFENODVPPERI 14037
QY 137 ----- 144
Db 14038 VSTLSAGSRDTSRNPVLVOLLFVAVHSSQOGLGRIQLDGVVDEPVLSTVSTRDLSEHAFQEA 14097
QY 145 ----- 144
Db 14098 DRLNGVMFATDLFQPETIOGVAVVEVLQGLEQPOSPIATMPLAEGIAQLRDAGALQ 14157
QY 145 ----- 144
Db 14158 MPKSDYPRNASLVDFVQOQNASPSTVAVTSDTSKLTVAELDRLSQOAAASYLRROQLPAE 14217
QY 145 ----- 144
Db 14218 TWAVLAPRSCETIIAFLAILKANLAYMPLDVNTFSPARMEAIISVPCRRLLLVGSGVRH 14277
QY 145 ----- 144
Db 14278 ADINVPNAKTMLISDTVTGTDAGTGPPLVVRPSATSLAYVFTSGTGKPKGVMEHRA 14337
QY 145 ----- 144
Db 14338 IMRLVKDSNVVTHMPATRMARHTNIAFDVSLFEMCATLLNGGTLVLCIDYLLTLDSTMRL 14397
QY 145 ----- 144
Db 14398 ETFEREQVRAAIFPPALLRQCLVNMPPDAIGMLEAVYVAGDFHRSRDARATQALAGPRYN 14457
QY 145 ----- 144
Db 14458 AYGPTENAILSTIYNIDKHPYNGVPIGSAVNSGAYVMDRNQQLPPGVGELWVTGE 14517
QY 145 ----- 144
Db 14518 GVARGYTDASLTDREFTVITDQQRQAVRTGDRVYRPRKPGQIEFFGRLDQQAQKIRGHR 14577
QY 145 ----- 144
Db 14578 VELGEVEHALLSENSVTDAAVLRITMEEDPOLVAFVTTDHEYRSGSSNEEDPYATQAA 14637
QY 145 ----- 144
Db 14638 GDMRKLRLSLLPYMVPVSRVTILROMPLNANGVKVDRKDLARRAQMTPTASSSGPVHVAPR 14697
QY 145 -----GVT 147
|:|

Db 14698 NETEAALCDEFETILGVKVGIT 14719
RESULT 7
Q93NW6 PRELIMINARY; PRT: 10917 AA.
ID Q93NW6
AC Q93NW6
DT 01-DEC-2001 (TremBurel. 19, Created)
DT 01-DEC-2001 (TremBurel. 19, Last sequence update)
DT 01-JUN-2002 (TremBurel. 21, Last annotation update)
DE AMPHC.
GN AMPHC.
OS Streptomyces nodosus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=40318;
RN [1]
RP SEQUENCE FROM N.A.
RA Caffrey P., Lynch S.V., Flood E.M., Finnian S.M., O'Liynk M.;
RT "The amphotericin biosynthetic gene cluster from Streptomyces
nodosus.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF357202; AAK73514.1;
DR InterPro; IPR001227; AC transferase.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR004410; Fabb.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantacn_attach.
DR Pfam; PF00698; Acyl_transf_6.
DR Pfam; PF00107; adh_zinc; 1.
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt_C; 6.
DR Pfam; PF00550; pp-binding; 6.
DR TIGRfam; TIGR00128; fabb; 6.
DR PROSITE; PS00075; ACP DOMAIN; 6.
DR PROSITE; PS00059; ADH_ZINC; UNKNOWN 1.
DR PROSITE; PS00606; B_KETOACYL SYNTHASE; UNKNOWN 6.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN 5.
SQ Phosphopantetheine.
KW SEQUENCE 10917 AA; 1132905 MW; 15AC5956BBS810A1 CRC64;
Query Match 31.4%; Score 243; DB 2; Length 10917;
Best Local Similarity 1.1%; Pred. No. 21;
Matches 94; Conservative 21; Mismatches 33; Indels 8751; Gaps 27;
QY 1 AAG----- 3
|:|
Db 384 AAGVKNMAIOHGVLPRLSHAERTPSNVDTAGAVELLDRAVDMPETGRARRAGVSFG 443
QY 4 ----- 3
Db 444 ISGTNAHLILEQAPTAPERPADPAARPALVPWTLSGRTAAALDAQRERLAHSHPGA 503
QY 4 ----- 3
Db 504 DPLDIGFALADGRATFEHRAVLACGTASAAATEIAHGVSVDGASAVLFGSGSORLGMG 563
QY 4 ----- 3
Db 564 RELHARFPVFAEFADEIAALLDRHTRPLREVVYVWGSDAELLNETGWTOPALFAVEVALYR 623
QY 4 -----GILHLE----- 9
|:|
Db 624 LVSSGLVTPDYVGHSIGEIAAAHAGVLSLEDACALVAARARLMQELPRGGAMLAVRAT 683
QY 10 ----- 9
Db 684 EEEVAPHLTERVSIASVNGPSSVVVAGDEDEVAIAAHFEAQDRKTTTLRLRVSHAFHSPLM 743
QY 10 ----- 9
Db 744 DPMLEEFRRVARGLTYHEPRIPVVSNTLGTAIADPADLCTADYVVRHVREAVRFADGVTTL 803

QY 10 -----LLVAVGPD----- 17
Db 804 ACRGVTTLLVGVDPGVLSSAMAQESLPEGTTGVPLLRKDRPEESALTGLRVHVRGLTVR 863
QY 18 ----- 17
Db 864 WARLFDSAGAGRADLPTYPFONRVMPAAGRSARTSPAAGLSADHPLLLGAAVELADGEG 923
QY 18 ----- 17
Db 924 FLFTSRLSVSHHPWLADHGVQAAPCVPGTAFVELAIRAGDEAGCDRVEBELTLAAPLVLP 983
QY 18 ----- 17
Db 984 RGAVALQVRVGAPDPAGRRTLTGTSERVEDGFDLPMSQHATGVLTTCAGAPDAADFATVMP 1043
QY 18 ----- 17
Db 1044 PAGAEQVDLDGCYERLAGLFDYGPAPRALRAAMRRGTEVFAEVALSENAGTDPADFGLIH 1103
QY 18 ----- 17
Db 1104 PALLDAAQHAAAYNDLGAISRGGLPFAMEGVHAAAGATTVRARITSVGEDTVSIAYDT 1163
QY 18 -----VFG----- 20
Db 1164 AGGTVLSVDLSVRAVPKDIPTCARAVHRDSLFOVDWTPLRQSQGLPETVALVGADPAG 1223
QY 21 ----- 20
Db 1224 LADTLRTGCVQTVHPDLAALADAPVPVLTALPGDQDTPAADAHTTAAAVLGLAQ 1283
QY 21 -----AHQEDTERYVLTNL----- 34
Db 1284 HMLTEDRFADARLVLTTRDATAGDLAAAAGGLIRTARTENPGRFVLLDLATGTPAAPA 1343
QY 35 ----- 34
Db 1344 ALVAALAAAYDEPDLAVRGSDVLTARLARPAPDAGADWNPDGTVLTGGTGLGVGLAR 1403
QY 35 -----VFG----- 34
Db 1404 HLVTERGVRLLLASRRGPAADGAPALVDELTALGASVDVAACDVTDREALAALLARVPA 1463
QY 35 ----- 34
Db 1464 AHPLSAVVHTAGVDDGVIGSLTAERLOTVLRPKADAAWHLHELTRDLDLDAFVLFSSVA 1523
QY 35 ----- 34
Db 1524 ATLGSPGOGNYAAGNAFLDALAAHRHALGLPATSLAWGPWTQSVGMTGLSDIDVERIAR 1583
QY 35 ----- 34
Db 1584 SCMPPLTVEQVALFDAALATGPPALLPVRLDLAVLRTQGEIAPLLRGLIRTTVRRAAQ 1643
QY 35 ----- 34
Db 1644 VSVTADGLAQOAGLADAAAREALLGLVRTQIAVLGHADAGQVEAARQFODLGFDSLTA 1703
QY 35 ----- 34
Db 1704 VELRNLNSATGLRLPATWVFDPYTPNALADHLSDELLETGTONIPLPTPAGALPATDDP 1763
QY 35 ----- 34
Db 1764 IIVIGMACRYPGCVTSPEDLWRLVSEGADATGFPFVNRGWDVNLVDPPDPRGHTYVRA 1823
QY 35 ----- 34
Db 1824 GGFLHDAGSFDADFFGMSPREALSTDSOORLLELSWEAVERAGIDPHSLRGSRTGVFAG 1883

QY 35 ----- 34
Db 1884 VYNDYGTILTGEYEAFRNGSAPSVASGRVSYTLGLEGPAVTVDTACSSSLVGMHLAA 1943
QY 35 ----- 34
Db 1944 QALRTGECTLALAGGVTVMTSTPSTFVDFSRORGLAADGRSKAFAPAAADVAMSEGVGMILL 2003
QY 35 ----- 34
Db 2004 LERRSDAIRNGHEILAVLRGSVNDGASNGLTAPNGPSQQRVIRQALASGGLSPEDVDV 2063
QY 35 -----NIG----- 37
Db 2064 VEAHGTGTTLGDPIEAQALIAIATYGRDRDPERPLWLGSVKSNIGHTQAAAAGVAGVIKIMS 2123
QY 38 ----- 37
Db 2124 MRHGLLPOTLHVDAPSSHVDWAGAVELLTEOTEPENDRVRRRAVVSFGISGTTNAHVII 2183
QY 38 -----AELLRD----- 43
Db 2184 EQPAAGPAPAPSPGAEELPAAPHPPLSAKTPEALRDQARLLAHVEAHPETRPADIAVSL 2243
QY 44 ----- 43
Db 2244 LTRSVPDHRRAAVLGADRTEALRALAAGTDDALLTGTARGORTAFLFSGOGSORLG 2303
QY 44 ----- 43
Db 2304 MGRBLYGRVPVTEALDAVLAVLDGELEGSUREVMGEDADLLNETGTQTPALFAVEVAL 2363
QY 44 ----- 43
Db 2364 YRLVESWCVKPDFVAGHSIGIEIAAHITAGVLTADACRLVAARAALMOALPEGGMIAVE 2423
QY 44 ----- 43
Db 2424 ATEDEVLLPLTGDVSIAAVNGPSSVVVSGAEAAALAVAGHFTGCGRTSRLRVSHAFSP 2483
QY 44 ----- 43
Db 2484 LMDPMLDEFRSVAEGLSYSAPAI PWVSNLTGTLADPADLCSADYVVRHVRDAVRPADCVG 2543
QY 44 ----- 43
Db 2544 TLGEHGVTTFVELGPDGVLSSMAQOOSVTEGAVTVPLRKORSEQALVTALCRLVVLGID 2603
QY 44 ----- 43
Db 2604 ADWTSALTVGGAARVDLPTVAFQHRWFMPAARAAHPDMRAAGLSASHPLLLGAAVELPG 2663
QY 44 ----- 43
Db 2664 GEGVLLTGRLSLQSHPLADHAVHGTVLLPGTALLDLAIRAGDEAGCGHLEDLTAAPLT 2723
QY 44 ----- 43
Db 2724 LPQDGA VLLQVRGAADDAGRRVTTVHARPEDATDGVWTLHATGTLSDTAATGALTDTTV 2783
QY 44 -----PSLGAQ-----FR 51
Db 2784 WPPAGAQAALDTEGCVQRFPAHGFAYGPLFQGLTAANKAGEALYAEIALPESGHDDAAGFA 2843
QY 52 ----- 53
Db 2844 VHPALLDAAALHPALLTDDGGLPFSWENSVLHAGATALRVRLAPAGANALSVTAADPAG 2903
QY 54 ----- 53
Db 2904 RPVVSIALRLARPVEADOLTDGACHLRDALFRLEWTPVLPDGTPTESLALLGPDTDLLA 2963
QY 54 ----- 53

Db	2964	EALGEGAVRYATLADLLAADAPVPGAVLVPLTAGSDTDDTPGAVHRTAAALTLLQEWLA	3023	Db	4044	GRELYGRFPVFAEALDAVLAVLGELEGSLREVMWGEDAGLLNSETGWTQPALFAVEVALY	4103
Qy	54	-----	53	Qy	68	-----	67
Db	3024	EDRFNGSRLVLRGAVATDDAPVTDLAAAALVGLVRSQAQENPGVFGFMDLDADGASDS	3083	Db	4104	RLVESWGVPRDPFVAGHSIGEIAAAHAGVFSLEDAARLVAARGRLMOALPAGGAMVAQOA	4163
Qy	54	-----	53	Qy	68	-----	67
Db	3084	ALTRALCLDEPQALRGRALAAARLTRATADSTARVAMSSEGTVLVTGGTGGGLGLVARH	3143	Db	4164	TEDEVIPYLSDEVSI AALNGPASVVVSGAEGAVLEVAARFEAEGRKATMLRVSHAFHSPL	4223
Qy	54	-----	53	Qy	68	-----NITANLTSLS-----	77
Db	3144	LVMERGVRRLLLTSRSLGAAGAGDLVAELEGAEVSVAAACDVADRDVATLVAAPAE	3203	Db	4224	MDLMDPEPRAVETLSFAAPVIPPVSNLTGSLATAEELCSPEYVVRHVREAVRFADGUSA	4283
Qy	54	-----	53	Qy	78	-----	77
Db	3204	HPLVAVVHTAGVLDGVLDSLTEERLATVLRPKVDGNWNLHEATRLDLDAFVVVFSSVAG	3263	Db	4284	LEAQGVTTTFLGPDGVL SAMAQESLTGDEAVTVALLRKORDEETAALTAALHHTAGVR	4343
Qy	54	-----	53	Qy	78	-----	77
Db	3264	VFGAGQANYAANAFLDALTAAHRLGLPAVSLAWPMQOGAGMTAGLDERDIRRAES	3323	Db	4344	LDWNAFFAGTGATRVLDLPTYAFQHTTYWPSGTHSADAAAAGVLTAAAGHPLLNGSVOLAEG	4403
Qy	54	-----	53	Qy	78	-----	77
Db	3324	GMPLITVEOGLAFDAALTDEAALVTSRLDLAAFRGRPEIPALLRGLIRTPVRRAAATG	3383	Db	4404	EGVLTGRLSLQSHPLADHAVHGTVLLPGTALLDLAPRAGDEVCCDRVEELTLAAPLVL	4463
Qy	54	-----	53	Qy	78	-----	77
Db	3384	AADGADAGLVORLAGLDREORHEVLLDLVRTSAALVLGHAAGSDIDADRAFRLGDFDSLT	3443	Db	4464	PERGAVQTVQVRGVADDTGRRRTTVVHSRPEDATDVSTQWTHATGTLMPGTGRTDGTGDTATV	4523
Qy	54	-----	53	Qy	78	-----	77
Db	3444	AVELNRLRTATGURLSATMVFDHPTLSALADHLLTFLFGAEATLTPVLLPPTDDPI	3503	Db	4524	WPPTGAELDTBGCYERFTALGFGYGPVFGQLKAAWRSQDVLVYAEVLOEAGAADAEGF	4583
Qy	54	-----	53	Qy	78	-----	78
Db	3504	VIVGACRFPGDVTSPEDLWFLVSEGTDAISDFPANRGNVDVSLYHPDPHSGTSYTRSG	3563	Db	4584	GLHPALLDAALHASLVAHEGEGNGGLPFAWEGAVLHATGATALRVLTPTDQARQSVAI	4643
Qy	54	-----	53	Qy	79	SVC-----GM-----	83
Db	3564	GFLHDAGDFDPDFGMSPREALATDSOQRLLETSWEAIERAGIDPLSLRDSRTGVFAGV	3623	Db	4644	AVCDTAGQPVASIDRLLVRAVTDGDLNEAAGIARDALFALGWTPLPQDGDIPAAVALVGP	4703
Qy	54	-----	53	Qy	84	-----	83
Db	3624	MVHDYSLGKGKEFGFGOGSAGSVASGRVSYTFGEGPAVTVDTACSSSLVALHWAQ	3683	Db	4704	DTTGLAADLTAAAGAAVTAHQDLAALAADGTDIPGTVVTVDTGTAAPVABAAHTVAAEALA	4763
Qy	54	-----	53	Qy	84	-----	83
Db	3684	ALRSGECSLALAGGVTVNMSTPTFIEFSRQOGLSADGRSKAFASADGVGHSEGVGMILL	3743	Db	4764	LAORWTAEDRFAGARLVFATCGAVEAGGSQVTDVAAAAMVGLVRSQAQSESPDTFVLVDRE	4823
Qy	54	-----	53	Qy	84	-----	83
Db	3744	ERQSDAIRNGHEILAVVRGSAVNQDASNGLTAPNGSPQORVIRQALASGGLSTGDVDAV	3803	Db	4824	AADAAGDAABERGSLFLAALGSGEPQLALRDDVVVLGRLARFADAAGANTLPAERAWHLDST	4883
Qy	54	-----LVKM-----	57	Qy	84	-----	83
Db	3804	EAHGTGTTGLDPIEAQALIATYQONRTDEQPLWGLSKLSNIGHTQOAAAGVAGVKNWMSM	3863	Db	4884	TPGSINGLALTPHPEALEPLTGHQARI EVRAAGLNFRDVLKALGNYPGDAGRYGHEAAGV	4943
Qy	58	-----VILT 61	61	Qy	84	-----	83
Db	3864	RHGVLPTLUVDAPSSHVDWTEGAVELLTEQTEWPETDVRREAAVSFFGISGTNAHVILE	3923	Db	4944	VVEGPDVTHIAPGDRVMGMVSGSPASPAVTDARRLTHLPEOCTWEIGASLPLVFLTAYH	5003
Qy	62	EPEGAP-----	67	Qy	84	-----	83
Db	3924	QPVAPAPATADVGAEPVPLILSGKSPALRQOARLLDVTREHTALRDLGHSLA	3983	Db	5004	ALKELGGLTAGKVLIIHAGAGVGMAAIOIARHFGAEVFATASEAKMDVLRSLRGVADDDHI	5063
Qy	68	-----	67	Qy	84	---SQT1-----	87
Db	3984	TSRSAPDHRVAVLATGREDALRALTAADDEANSAAVTGTRSGRRAALFSGQGSQRLGM	4043	Db	5064	ASSRTLDFEAAFAAVAGENGLDVVLNSLAGEFVDASHMLLTGTGGRFLEMCKTDIREDDSV	5123
Qy	68	-----	67	Qy	88	-----	87
Db	5124	PDGITYOSFDLAFVDPEVIGMNGRELTTELFAAGDLHPLPVRAWMDIRHATDAFRHMSMARH	5183	Db	5124	PDGITYOSFDLAFVDPEVIGMNGRELTTELFAAGDLHPLPVRAWMDIRHATDAFRHMSMARH	5183

```
QY 88 ----- 87
Db 5184 IGKIVLTPRAWSEGTVLVTGGTGGGLVARHLWMERGVRRLLLTSRSGLDAGAREL 5243
QY 88 ----- 87
Db 5244 VAELENLGAESVAACVADRDAVDTLIAGIPAELPRVHVHTAGVLDGVLGSLTEERL 5303
QY 88 ----- 87
Db 5304 ATVLRPKVDGAWNLHEATRRLLDAFVVFSSVAGVFGGAGQANYAAGNAFLDALTAHRQS 5363
QY 88 ----- 87
Db 5364 LGLPGLSLAWGAWAOGAGTAGLBERDIRRAAESGMPLITVEQGLALLDALTGSAALV 5423
QY 88 ----- 87
Db 5424 PVRDLAVLRARGTVAPLMRGLVRAPARPAATAATGDTALVDRLTRLORTERRDALLTL 5483
QY 88 ----- 87
Db 5484 VREQAALVLHGSGGGIDPSRAFRDLGFDLSLTAVELNRNLGAATGVRLSATAVFDPYTV 5543
QY 88 ----- 91
Db 5544 ALVDHLLTELLGPDAESDAEAPAEALLDDDPPIVIGMACRFPGGITSPEDLWRLLEG 5603
QY 92 ----- 98
Db 5604 SDVISDFPNRGWDVDSLYDPAPHTGTSYTRSGGFLHNAADFADDFGMSPREAMATDS 5663
QY 99 ----- 98
Db 5664 QORLLESTWEIERAGIDPVLGRSGTVGFAGVMYSGYGTALDGDDEFEGQGSALS 5723
QY 99 ----- 98
Db 5724 ASGRVSYTFEGEPAMTIDTACSSALVAMHLAAQALRSCECSLALAGGTVVMSIPDTFVE 5783
QY 99 ----- 98
Db 5784 FSRQGLSPDRCKPFSESADGVHSEGVGMULLLERQSDAIRNGHRVLAVVRGSAVNQDG 5843
QY 99 ----- 98
Db 5844 ASNGLTAPNGPSQORVIRQALASGGLSTGVDVAEAGHTGTTGLDPIEAQALLATYQNR 5903
QY 99 ----- 98
Db 5904 PEDRPLLGSVKSNIGHTQAAGAAGVIKVMAMQHGVLPRSLNISEPSSHVDMSAGAVE 5963
QY 99 ----- 98
Db 5964 LLTEQTEWETDVRVRAGVSSFGISGTNAHVILEQPAKVIQGTVAASAPGDPVVEPAV 6023
QY 99 ----- 98
Db 6024 VPWVLSKGTPEALRDQAARLLATVTEEPAPRPVDVALSLATQRSQFAHRGVVLAGEHEET 6083
QY 99 ----- 98
Db 6084 VRLAALADGEPHSSVTGTGVSSGRRALFSGOGSRLGMGRELXGRFPVFAELDAVLA 6143
QY 99 ----- 98
Db 6144 VLDGELGSLREVMWGEDAGLLNETGHTQPALFAVEVALYRLVESNGVRPFDVAGHSIGE 6203
QY 99 ----- 98
Db 6204 IAAAHAGVFSLEDAARLVAARGRLMOALPAGGAMVAVQATEDEVVPHLSGDVSIALLNG 6263
```

```
QY 99 -----DLVLYI--TRFDLE----- 110
Db 6264 PTVSVVSGVEDAVLEVAARFEAEGRKTRURVSHAFHSPHLMDEPMLDEFRAVAGLSFAAP 6323
QY 111 ----- 110
Db 6324 VIPVSNLTGSLATAEELCSPEYVVRHVREAVRFADGVAALAEAOGVTTFLELPGDGVLSA 6383
QY 111 ----- 110
Db 6384 MAQESLTGDETVTIPLLRKORDEETAALTALGRLLHATGTRVDWAGFLAPAGARTVDLPTY 6443
QY 111 ----- 110
Db 6444 AFQHRPFWPSGKRGATDVNSVGITAAAHPLLNVAVELAEGEGLVFTGRLSUNSHWLADH 6503
QY 111 ----- 110
Db 6504 AVMTGVLLPGTALLELALRAGDEVGCDHVBELTLAAPLVLTEOGAVRTOVRVGAEDAAGR 6563
QY 111 ----- 114
Db 6564 RTVTIHSRPERSTDAPWTHQATGVLATGLPTNAAPPDATVMPADAEFVDLTGTFYEAREA 6623
QY 111 -----LPDG----- 114
Db 6624 EGPAYGOAFOGLKAAMRRGDEVFAEITLPDGTOKEAESFGLHPALLDAGLHAAMLAAPT 6683
QY 115 ----- 114
Db 6684 DGTGTVPFWMGVSLVSGASSVVRVLRGRTDGTTLTAVADTTGAPVAAVQALTMRAVS 6743
QY 115 ----- 114
Db 6744 AEALSSAALVRDSLFRLDWTPVTPAGSAPTAALVGEDPGLVRLREAGTEIGRRASL 6803
QY 115 -----NRQV----- 118
Db 6804 TDDAELPDLILLPVTAEEHEEAADTPAVTHALTQVLELVRARLEDERHAATREFVFTRG 6863
QY 119 ----- 118
Db 6864 ATTGHDLAAAVWGLVRSASENPGCFGLIDLDPETTRPLPLTALLGDEPOLRLODDLR 6923
QY 119 -----RGVTQL----- 124
Db 6924 AARLVRPAPATDAVSAGGEGAVLVTCGTGGLCAVLARHLVAEHGVRVLSRRGAA 6983
QY 125 ----- 124
Db 6984 AGAAELVAELAESGARATVACDVTDRAAVAELVAAPVSAVHVSAGVLDGLIASLTPE 7043
QY 125 ----- 124
Db 7044 RLSGLVRPKVDAANLHEATRDLDLDAFVVFSSVAGVFGNAGQANYAAGNAFLDALMEHR 7103
QY 125 ----- 124
Db 7104 RAAGLPGLSLANGPWEQTGCGMTGTLSDAELERLASSGVPPLAVEOGLALFDALSTED 7163
QY 125 ----- 124
Db 7164 AALVPVRLDLPALRGQGEVPPLLRSLIRASRRRAAVAGSATATGLRDLRLAVLDAVERREV 7223
QY 125 ----- 124
Db 7224 LVDLVRGQVALVGHADATAVHPTRAFRDLGFDLSLVELNRNLNAVTLGLRLPATWFDY 7283
QY 125 ----- 124
Db 7284 PTVEVLVFIDELGLGTEVEAEVVRGTAAVDDPFIIVGMACRYPGGVASPEDLWRLVT 7343
QY 125 ----- 124
```

Db 7344 EGTDAVSAPPVNRGMDVENLYHPDPDNIGTSYTRSGGLHEAGEPDPFGMSPREALST 7403
Qy 125 ----- 124
Db 7404 DSOORLLLETSWEATERAGIDPVGLRGSAATGVFAGVMYSAMLASPEFEGFQSGSSP 7463
Qy 125 ----- 124
Db 7464 SVASGRVSYTFGEPGPAVTDTACSSSLVAMHWAMQALRSGBISLALAGGVTVMSTPGVF 7523
Qy 125 ----- 124
Db 7524 VDFARORGLSPDRCKAFSDSADGVGMSEGVGMVLVLEROSDAIRNGHEITLAVVRSAVNO 7583
Qy 125 ----- 124
Db 7584 DGASGLTAPNGFSQORVIRQALASGLSAQIEDVVEAHGTGTLGDPTEAQALLATYGR 7643
Qy 125 ----- 124
Db 7644 DRDPEQLMLGSVKSNIIGHTQAAAGVAGVIKVMMSMRHGVLPRTLHVDAPSSHVDWTEGA 7703
Qy 125 ----- 124
Db 7704 VELLTQTEWPETEHVRAGVSSFGISGTNAHVILEOPAKVIOGTVIGGSTPESGVVEPS 7763
Qy 125 ----- 124
Db 7764 VWPVLSGKTPEARLSQAOKLLASVEAELDRPLVDVGSSILVAARSLFEHRAVVLATDADT 7823
Qy 125 ----- 124
Db 7824 AARALAALVGEPPAAVSGPARTGRSAALFSGQSORLGMGRELYGRFPVFAALDAVL 7883
Qy 125 ----- 124
Db 7884 AVLDELGDSLRVEMMGEDAGLLNETGWTQPALPANEVALHRLVSEFVTPDFVAGHSIG 7943
Qy 125 -----GCA----- 127
Db 7944 EIAAAHTAGVFSLEDAARLVAAARGRLMQALPAGGAMVAQATEDEVIPYLSDEVSTAAIN 8003
Qy 128 ----- 127
Db 8004 GPASVVYSGAEGAVLEVAARFEAEGRKATMLRVSHAFSPMLDPMLEDFRAVAEGLSFAA 8063
Qy 128 -----CSP----- 130
Db 8064 PVI PWVSNLTGSLATABELCSPEYVVRHREAVRFADGVTTLEEQVTTLELPGDGVLS 8123
Qy 131 -----TWSCLE----- 135
Db 8124 AMAQESLTGDEAVTVPLLRKNTAEAAATVAALARLHIGGLRIDMSALFAGTGARRVELPT 8183
Qy 136 ----- 135
Db 8184 YAFQHWFWPAAPLGGDVRAAGLSAEHPLLGAAVELAAGEGVLTGRLALHSHPLWAD 8243
Qy 136 ----- 135
Db 8244 HTVDGTULLPGTALLELAIRAGDEVGSDQVEELTITAPLVLPERGAVQVQVAEPDAAG 8303
Qy 136 ----- 135
Db 8304 RRVGVYARPECTGTANSQHAAGVLAAGSAPAPAPADTAVMPPEGAQAVDADGCVEEF 8363
Qy 136 ----- 135
Db 8364 AAAFGYGYPLFOGLRAAARRGDEWFAEVALPEGTTAAGFGLHPALFDSALHALLLTPA 8423
Qy 136 ----- 135

Db 8424 ADGCTAGLPESWTDVSLHUSGATVLRVRLTPTGDRTLSLSAVDPGTPGVVSGSLVTRSV 8483
Qy 136 ----- 135
Db 8484 DGLTTCTEALLRDADFEGVEWTPAQGSEAAATVALLGTGDDPLGLAAALTGAGTTVRTY 8543
Qy 136 ----- 135
Db 8544 PDVAALVAATVAADGPVPEVVVHGVASHDGEPAASDAARHLTADALDTVQSRLSEDRLAGA 8603
Qy 136 ----- 135
Db 8604 RLVEVTRGAVDTRDLGAAAVQGLVRSQOTENPGALGLVDLGTTPESAALAAALGSAEPQ 8663
Qy 136 ----- 135
Db 8664 LALRDGRPHARLARLARPADTAAPTAWQSDGTVLVTGGTGLCAQFARHLVDAYGVNRL 8723
Qy 136 ----- 135
Db 8724 LLVSRGGDAPGTTTELVAELIAHGAEVTVQACDVADPDVATVAVAGIPAEHPLTAVVHTA 8783
Qy 136 ----- 135
Db 8784 GVLDDGVIGSLTGERLARVLKADAAWNLHQATRGDLDDAFIVPSSVAGVFGGAGQANY 8843
Qy 136 -----ITE----- 138
Db 8844 AAGNAFLDALMEHRRACGLPGLSLAWGPMDOAGGTMETLSDAEARLARAGMPPLAABOG 8903
Qy 139 ----- 138
Db 8904 FALFDAAALAGNGTGTGSEDITGTALVVPVRLDLAALAAOGEVPAVLRLVTRTRRTVA 8963
Qy 139 -----DTGFDL-- 144
Db 8964 GGTVTADGLVARLTALTAEERREALGLVTRQAAVLVGHADAASVDAQAQFQDVGDSL 9023
Qy 145 ----- 144
Db 9024 AVELNRLSAATGLRLTATLVFDYPNAATLASHLHDELFGAETESALALPVKQLPATTTDD 9083
Qy 145 ----- 144
Db 9084 PIVIVMACRYPGVSPEDLWRLLEGTDAITEFPVNRGWDIESVYVHDPPEHLGTSYTR 9143
Qy 145 ----- 144
Db 9144 SGGFLHEAGEPDPFGFMSPREALATDSOORLLLESSMEAIERAGIDPVGLRGSATGVFA 9203
Qy 145 ----- 144
Db 9204 GVMYSYDGSILGKKEFEGLOGGSAGSVASGRVSYTFGEPGPAVTDTACSSSLVAMHILA 9263
Qy 145 -----GVTI 148
Db 9264 AQALRTGCSLALAGGTV 9282

RESULT 8
Q9JNX9 PRELIMINARY; PRT; 9510 AA.
ID Q93NX9;
AC Q93NX9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Amph.
GN AMPHI.
OS Streptomyces nodosus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=40318;
RN [1]

RP SEQUENCE FROM N.A.
 RA Caffrey P., Lynch S.V., Flood E.M., Finnan S.M., O'Liynyk M.:
 RT "The amphotericin biosynthetic gene cluster from Streptomyces
 RL nodosus";
 DR Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF357202; AAK73501.1; -;
 DR InterPro: IPR001227; Ac transferase.
 DR InterPro: IPR002198; ADH short.
 DR InterPro: IPR004410; FAbD.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR000794; Ketoacyl-synt.
 DR InterPro: IPR001005; Myb_DNA_binding.
 DR InterPro: IPR003880; Ppantne_attach.
 DR Pfam: PF00698; Acyl_transf_5.
 DR Pfam: PF00106; adh_short_1.
 DR Pfam: PF00109; ketoacyl-synt_6.
 DR Pfam: PF02801; ketoacyl-synt_C_6.
 DR Pfam: PF00550; pp-binding_6.
 DR TIGRFAMs: TIGR00128; fAbD; 6.
 DR PROSITE: PS50075; ACP_DOMAIN; 6.
 DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_6.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
 DR PROSITE: PS00012; PHOSPHOPANTHEINE; UNKNOWN_6.
 KW phosphopantetheine.
 SQ SEQUENCE 9510 AA; 992817 MW; 1A8E003A15D478CF CRC64;

Query Match 31.2%; Score 241; DB 2; Length 9510;
 Best Local Similarity 1.0%; Pred. No. 18;
 Matches 95; Conservative 18; Mismatches 31; Indels 9162; Gaps 25;

QY 3 GGILH----- 7
 DB 88 GGFLHDAPEFDADFGISPREAVAMPDQORVVLESANEAFERAGINPTSVKSGRTGVIG 147
 QY 8 ----- 7
 DB 148 AMAQYRVCPADAGFQLTGNTGVSLSGRISYTFGTGPAVTVDTACSSSLVAVHLATQ 207
 QY 8 ----- 7
 DB 208 ALRAGECTLALAGGVTVMSGPGTFIEMGRGGLSVDRGCRSFGDTADGTGWAEGVILVL 267
 QY 8 ----- 7
 DB 268 ERLSDAIRGREILAVVRGTAVNQDASNGLTAPNGPSQOAVIEQALYNARLSAGDIDV 327
 QY 8 ----- 7
 DB 328 EAHGTGTLGDPVEAQALLATYQQORDEKPLLLGSKVSNISHTQAAAGVAGVVKVMAM 387
 QY 8 ----- 7
 DB 388 RHGVPLRTLLADEPFRHVDMSGAVRVLTEWTPATGAPRAAVSSFGISGTNAHTIVE 447
 QY 8 ----- 7
 DB 448 QAPEPEPADPEDAPSTPAATGVLVPLLSGRSFEVLRAQAAALLTLTGTPPPAPADLA 507
 QY 8 ----- 7
 DB 508 YSLATRTAFEHRAVLLASDLPETLGRLTAAEGTDPVLAADVGTGTARTETTRAVLFTG 567
 QY 8 ----- 7
 DB 568 QGAORLGAGRELAARFPFAAALDAALDAFTPHLDVPLRKVLWGEDADRLDRTEYAQPAL 627
 QY 8 ----- 7
 DB 628 FAVEVALYRLLESFEVKPDHLACHSVGEIAAAHVAGVFSLLDDAATLVAARGRLMQALPEG 687
 QY 8 ----- 7

DB 688 GAMVAVOASEDEVAPLLACHEDLVSLAANGPSAVVLSGDETTVTTELAARLAADGRKTSR 747
 QY 8 ----- 7
 DB 748 LRVSHAFHSPMLAPMLDEFNNVVEGTLHSPLLPVVSDVTGEPATVAQLTSPDYVWDHVR 807
 QY 8 ----- 7
 DB 808 QAVRFADGIDMLARHDVTAFLGLSPDSVLSAMAQNCIDAAGSALTVPALREGRPEDHTF 867
 QY 8 ----- 7
 DB 868 TAALAAHLTGCTALHWDACFTGTGARRTDLPTVTFQRRRYWPRAVOGGAADLRSLVGLAA 927
 QY 8 -----LELLV----- 12
 DB 928 HHPLLSAAVSLADSEGALLTGRISSLSHPMLADHTVRGATLLPGTAPLELAVRAGDEVGC 987
 QY 13 ----- 12
 DB 988 DRVDELTLAAPLVLPQGGVQVOLWIGNPDASGRRSVTVYGRPDADADAPWTSHATGVLS 1047
 QY 13 ----- 12
 DB 1048 ASRTSDFDATVWPADAETLPVDGLYERLAEGGFGVPLFQGLRAAARRRGEVFAEWL 1107
 QY 13 ----- 12
 DB 1108 PESGHTDAESFGLHALLDSALHAASPFVLDERRAAGLPSFSEGLSHASGATTLRLVRLA 1167
 QY 13 ----- 12
 DB 1168 PAAGDAVAIAVADDGQGLVLSADSLILRAVAAREIDAANAALVRDALFRLDWVPVTAAS 1227
 QY 13 ----- 12
 DB 1228 GTAALVGEDPFGRLALPQFGDLAVHPDLADLAAADGAVPDTVLLPLTGTGPDADPTAA 1287
 QY 13 ----- 12
 DB 1288 HRAATEALAAVRTWLEQDERFAASRLALVTRGATTGHDPAANAAMVGLVRSQAENSENPRFL 1347
 QY 13 ----- 12
 DB 1348 LVDLDADQDTPALPAAALTSEEPQAVRGEBELRAARLVRRPASTAEAVPAGGEGAVLVT 1407
 QY 13 ----- 12
 DB 1408 GGTGGLGAVLARHLVAEHGVRELVLVSRRGGAAGAELVAELAESGARATVWACDVDR 1467
 QY 13 ----- 12
 DB 1468 AAVAEVAAHPVSAVWHSAGVLDDGMVGTLLTPERLTTVLRPKVDAANLHEATRDLDKA 1527
 QY 13 -----AVGP----- 16
 DB 1528 FVLFSVAGVLGSPQANYAAGNAPLDAALAAHRAAGLPGLSLANGPWEQTGGTGGISE 1587
 QY 17 ----- 16
 DB 1588 DDLRRMARAGTPALTVEOGLALLDALLDGDAAALAPVRLDLSVLRQAGEVPPLLRSLIRG 1647
 QY 17 ----- 16
 DB 1648 RSRRAAVAGSATAGLAQRLARLDAESRDELVLDIRGQVALVGHATGAIEIDAGRARE 1707
 QY 17 ----- 16
 DB 1708 LGFDSLTAVELNRNLNTVTGLRLPATLVFDYPTVSHLASVYLDLGLTTEVEAEVVQRCTA 1767
 QY 17 ----- 16
 DB 1768 AVADDPVIVMACRYPGGVTSPEDLWRLVTEGTDVAVSGFPVNRGMVDVNLVHPDPDHFG 1827

QY	17	-----16	30	-----29
DB	2908	DGEDQVALRATGVHRRLLVRHIVDELPSADQFTASGSVLIITGGTCGLGAETARWLARSGA	2908	DGEDQVALRATGVHRRLLVRHIVDELPSADQFTASGSVLIITGGTCGLGAETARWLARSGA
QY	30	-----29	30	-----29
DB	2968	AHLVLTSSRGDPAPGAELRAELBESGASVSIVACDVADRDLAAVLDCLSADQPLTGV	2968	AHLVLTSSRGDPAPGAELRAELBESGASVSIVACDVADRDLAAVLDCLSADQPLTGV
QY	30	-----29	30	-----29
DB	3028	HTAGVCHYGPLDALTPEAFAGLTAALAGAAHLDNLLGDRELDFFILFSGIAGVMGSGDQ	3028	HTAGVCHYGPLDALTPEAFAGLTAALAGAAHLDNLLGDRELDFFILFSGIAGVMGSGDQ
QY	30	-----29	30	-----29
DB	3088	SAYGAANAYLDALARRARRAATSIANGPMGCTGMAADDAVSGTLRRQGLGLDPPAPA	3088	SAYGAANAYLDALARRARRAATSIANGPMGCTGMAADDAVSGTLRRQGLGLDPPAPA
QY	30	-----29	30	-----29
DB	3148	LTEMRRAVVRQDVTTVADVDMTRYAPLFTSARPSALISDLPEVRALAAENTPADTGDAS	3148	LTEMRRAVVRQDVTTVADVDMTRYAPLFTSARPSALISDLPEVRALAAENTPADTGDAS
QY	30	-----34	30	-----34
DB	3208	EIVQVRSLSEPEQLRLTDLVRTEAATVLGHSSAGAVPEDRAFREIGFDSLTAVELRXH	3208	EIVQVRSLSEPEQLRLTDLVRTEAATVLGHSSAGAVPEDRAFREIGFDSLTAVELRXH
QY	35	-----34	35	-----34
DB	3268	LGAATGLSLPSTMVFDYPTPLEAQYLRAEMVGSVLEVAGPVATGTTDDEPIATIGMSCR	3268	LGAATGLSLPSTMVFDYPTPLEAQYLRAEMVGSVLEVAGPVATGTTDDEPIATIGMSCR
QY	35	-----34	35	-----34
DB	3328	YPGVSSPEQLWDLVLSGTDAITDPPVNRGWNAGLYDDPDHPGTYTSTQSGFLHEAD	3328	YPGVSSPEQLWDLVLSGTDAITDPPVNRGWNAGLYDDPDHPGTYTSTQSGFLHEAD
QY	35	-----34	35	-----34
DB	3388	FDPMFEGISPREALVMDPQORLLETTWEAFERAGLTPTDLRGLTGTFTIGSSYQBYGMG	3388	FDPMFEGISPREALVMDPQORLLETTWEAFERAGLTPTDLRGLTGTFTIGSSYQBYGMG
QY	35	-----34	35	-----34
DB	3448	AGDAGEHLVTGTSPSVLSGLAYVFCLEGPVTVDTACSSSLVALHACOALRNGESNL	3448	AGDAGEHLVTGTSPSVLSGLAYVFCLEGPVTVDTACSSSLVALHACOALRNGESNL
QY	35	-----34	35	-----34
DB	3508	AVAGGATVMTPNAFVAFSRQALAQDGRCKAFSESADGMTLAEGVIVLVERLSDARRN	3508	AVAGGATVMTPNAFVAFSRQALAQDGRCKAFSESADGMTLAEGVIVLVERLSDARRN
QY	35	-----34	35	-----34
DB	3568	GHPVLVIRGSAINQDGSNGLSAPNGPSQORVIRQALANARVAPGEIDLLEAHGTGTPL	3568	GHPVLVIRGSAINQDGSNGLSAPNGPSQORVIRQALANARVAPGEIDLLEAHGTGTPL
QY	35	-----37	35	-----37
DB	3628	GDPIEAQALFATYGRTRTPETALLGSKVNSIGHSSAAGVASIHKVMALRHGVMPOTL	3628	GDPIEAQALFATYGRTRTPETALLGSKVNSIGHSSAAGVASIHKVMALRHGVMPOTL
QY	38	-----37	38	-----37
DB	3688	HADEFSSHVDWSPGTVLLGENTDWPOTGRPRRAVSSFGISGTNAHVILLEQETEAFAAE	3688	HADEFSSHVDWSPGTVLLGENTDWPOTGRPRRAVSSFGISGTNAHVILLEQETEAFAAE
QY	38	-----37	38	-----37
DB	3748	DEQLAPAPLPVAAGVVPMLLSARGAAALREQADRLTLHVTADPAARPIDIGLSLATSRA	3748	DEQLAPAPLPVAAGVVPMLLSARGAAALREQADRLTLHVTADPAARPIDIGLSLATSRA
QY	38	-----37	38	-----37
DB	3808	LFEHRAVVVPAGTDPLEALRAVAADGPGSVVARGVADVAGRTVVFPGGSGOWAGMCAQ	3808	LFEHRAVVVPAGTDPLEALRAVAADGPGSVVARGVADVAGRTVVFPGGSGOWAGMCAQ
QY	38	-----37	38	-----37
DB	3868	LLDESPVFAERIAECAAALAEFTDWNLLIDVLRGAEGAPTLEVDVQVQSPAFVMSLAUV	3868	LLDESPVFAERIAECAAALAEFTDWNLLIDVLRGAEGAPTLEVDVQVQSPAFVMSLAUV
QY	38	-----37	38	-----37
DB	3928	WRAQGVPEPDVVGHSSQGEIAAAVVSGLSLRDGARVVTLRAQAIGRSLAGRGMMWSVALP	3928	WRAQGVPEPDVVGHSSQGEIAAAVVSGLSLRDGARVVTLRAQAIGRSLAGRGMMWSVALP
QY	38	-----37	38	-----37

Db	3988	VAEVEARLEAFEGRVSVAAENGRSSVVAGEPEALDELHAQLTAAEIRARRVAVDYASHS	4047
Qy	38	-----	37
Db	4048	PHVEDLHDEILLELAEVAPRTSEIPFFSTVTGDMWDTTMDAGYWRSLRGVLFADAVR	4107
Qy	38	-----	37
Db	4108	DLIAADHRAFIENVSSHPVLAMSVQMDIDDAGVAGVASCGLTRDNGGLDRFLLSAAEVFVR	4167
Qy	38	-----	37
Db	4168	GVQVDAAVFEFGTASRVLDLPTAFQHENLWAMAAAPEAVTAADPEDAAFTAVEDEGDVS	4227
Qy	38	-----	37
Db	4228	ALTAALGTDEDSVAAPALPSSWRRARKERSTVDSWRYRPTWKPVTKLPQRTLDGTWLLV	4287
Qy	38	-----	37
Db	4288	SADGVDDTDVAEALETGAEVRRLLVDESCTDRAVLRLRLTDADGLTGIVSVLAGAERTG	4347
Qy	38	-----	37
Db	4348	AVPGTGLVGLVALTVALVQALGDAGIDTPLWALTRGAVSTGRSDKVTPAPVQAVTGIGWT	4407
Qy	38	-----	37
Db	4408	AALECPGRGWGVVDPETILDARAGORLAVALAGALGDDDDQIALRSSGVFTRRIVRADAAP	4467
Qy	38	-----	37
Db	4468	DCSARDWKPRGTTLVGSGGTLAPHLARWLAEQGAHVLVLSRRGPEAPGAELRAELAE	4527
Qy	38	-----	37
Db	4528	RGTEITLAACDITDRDAVAALLLESUKAEGRTVRTVVHTAATIELHTLDATLDDDFRVL	4587
Qy	38	-----	37
Db	4588	AKVTGAQILDELLDDELDFFLYSSTAGMGGGAHAAYVAGNAYLAALAEHRRARGUTA	4647
Qy	38	-----	37
Db	4648	LSLSWGIWADDLQGRVDPQMIRRSGLFEMDPQLALSGLKRALDDDEQVIAVADVDMETY	4707
Qy	38	-----	37
Db	4708	HPVYTSARPTPLFDEVPEVQRLTAAAEQSAGDPARGFAAALLALPAAEQHRKLLTETVRT	4767
Qy	38	-----	37
Db	4768	EAASVLGLSSABDLTDQRAFDRVGFDSLTVAGLRNRLASVTGLTLPSTWTFYPNPAALA	4827
Qy	38	-----	37
Db	4828	GFLHSELADVHSAGAVATAGAPVDDDDPIAIVGMSCRYPGGITSAEQLWRVSLVEEDAVS	4887
Qy	38	-----	45
Db	4888	VFPADRGWDAEALYDPPDASGRYSVQGGFLRDVADFPDGFPGFISPREALSMDPQORLL	4947
Qy	46	-----	51
Db	4948	LETAWEFENAGLDPVAORSGRTGTIGASYYQDYGAAPVSGSEGCHMITGSLPSVLSGR	5007
Qy	52	-----	54
Db	5008	VSYLFCEGPVNTLDTACSSSLVAIHLACQSLNGESTLALAGGASINSTPMISFIFGSRQ	5067
Qy	55	-----	54
Db	5068	RALLAEDGRCKAYBAGADGMTLAEVGLILLERLSDARRNGHEVLAVIRGSVNVQDASNG	5127
Qy	55	-----	54
Db	5128	LTPNGPSQORVTRQALANAGVEANDIDVLEGHGTGTALGDPIEAQALFATYCKDRDPER	5187
Qy	55	-----	62
Db	5188	PVLLGSVKSNIGHTOMASGVASIIKLVLHALREGVAPKSLHIDOPSTHVDWSSGTIQLLTE	5247
Qy	63	-----	62
Db	5248	RTEWPTGRPRRAAVSSFGLSGNTVHTVLEQAPAADAPAAEDTPAPRDALVPVLVSRGE	5307
Qy	63	-----	62
Db	5308	AALRAQAGALLDLLAERPGIHPTDLAFSLATSRAALEHRAAVVADDHEALVRLGTALROG	5367
Qy	63	-----	62
Db	5368	LPGAGLVQGRTRGRGRTAFLTGQGSORLGMGRELYERHPVFADALDAVLARIDGTTTERPL	5427
Qy	63	-----	62
Db	5428	RDVLFRAEGSDAALLHRTGYAQPALFALEVALFRLLSHSGVTPDYLACHSVGEIAAAHV	5487
Qy	63	-----	65
Db	5488	AGVLDLDDACTLVAAARGRLWQALPEGAMVYLAEEDEVLPHEGLADQVVAANGVPRS	5547
Qy	66	-----	69
Db	5548	VVAGEBEPVLAAAHFAEOGRRTKRLRVSHAFSPMLDPMDDFAAVARALTYHAPSIP	5607
Qy	70	-----	69
Db	5608	FVSNVTGLAAPEOVCTADYVWSHVRSVRFADGIGHLSTGGVQVTFLELPGDGLSGMA	5667
Qy	70	-----	69
Db	5668	RESLTDASRTALLPTLRGDRPEEQALVTAVAHAHAGFDVDTAWFOGSGARRVALPTYA	5727
Qy	70	-----	77
Db	5728	QFERYWPDTTAAGITAPAPGSALDAEFMAAVEHADVASLTSLSGLDDATVTAMVPALTA	5787
Qy	78	-----	77
Db	5788	WRORRQSQSALDSWRYRVTKPRGGAPGAAPTGRWLVLPFAEHRDEATAAADAADVEAALA	5847
Qy	78	-----	77
Db	5848	TATVREVTGTDRAALAARLTAADGDTFQGVLSLLALAPGDAGHPCGAPALTLTATALQ	5907
Qy	78	-----	77
Db	5908	ALGDARIDAPLWNITRGAVAVGRSEQVTAPEQAAVWGLFRAAALELPAVGGSVDLPEDL	5967
Qy	78	-----	77
Db	5968	DTQAARLRGILAAADGEDAVAVRASGVFLRLLAHSPAADTVGSFDPNAGTVLITGTG	6027
Qy	78	-----	77
Db	6028	GIGGHLARRLARDCAAHLLTSSRRGPDPAPGAGELRAELEESGARVTIAACDAADRDLAA	6087
Qy	78	-----	77
Db	6088	LLATVPEDAPLTAVFHTAGVVDHVDDELTPESFATVLHAKTVAAARHLHDLTAGHDLAA	6147
Qy	78	-----	77
Db	6148	VLFSSTAGLGAAGQGNVAAANSQDLDALAEORRAOGLPALSVANGPWAGSGMWDAAEIA	6207

QY	78	-----	77	QY	86	-----	85
Db	6208	ARVRGGPEPLDPASGAALLRAVDNGDT\$VAADIDWERFLRAFASARPLPLVADLPET	6267	Db	7288	TASGLDGDVTVTMVPAISAWRRRTDARSTARDWRYHESMTALCTPAHSAGAGRVLALVP	7347
QY	78	-----	77	QY	86	-----	85
Db	6268	AGANAPAAAAGTGSGUREQIAGLPABERHAHVLDLLRTQVAAVLGHSRDARTVDDDLAFRD	6327	Db	7348	AEHAGTDWAETLVAALGADPLVVDGTSGLARELADLVPOQATFAPSRRTPARTLAAPAES	7407
QY	78	-----	77	QY	86	-----	85
Db	6328	LGFD\$TLVLELRNALNLTGLSLPASLVYDHTPREMADFLAELLGSLPESTRHAVATR	6387	Db	7408	PSTSTGTSGRHTESTHRTPLLDGQTS\$AGTLD\$AGT\$VTVVSLAAGTGLTLPADAPAA	7467
QY	78	-----	77	QY	86	-----	85
Db	6388	AVDEDP\$IAVVGLACRPF\$GVSTPEELWQLLAEGRDGITGFPDRNWDIAALGAGASDTLH	6447	Db	7468	VLEALEAAGVDAPLWCVTRGAVSVAGEAPAAVGOAALWGMGRVAALEHPERF\$GLADLAP	7527
QY	78	-----	77	QY	86	-----	85
Db	6448	GGFLAQVADFARFFGISPREALAMPQORLLLETTWEALERAGIDPAALRGSTTGVFVG	6507	Db	7528	DADAATAALLAHLAEPGGEDQIAVRATGLFRRLVRTAVAPGDSGMRPHGT\$VTVVVGTR	7587
QY	78	-----	77	QY	86	-----	92
Db	6508	TNGQDYNLLRRSTD\$GVYATGNTASVMSGRLSYALGLEGPAVTIDTACSSSLVALHW	6567	Db	7588	AMGARAARWLAREGAARLVLT\$TSPADSATDTEELRAELGRGA\$EVTVAPYDGGDRDAARA	7647
QY	78	-----	77	QY	93	-----	92
Db	6568	AGHALSAGECDLVVAGGVSVMSTPDSFVFPSTQGLAPDGHCKP\$DDADGTAMSEGVGI	6627	Db	7648	LLDGLTGLTAVVYADDT\$PADGPA\$AALAPVDTLAEAAATGRSLDAFVLFGSVAGVMGVRGT	7707
QY	78	-----	77	QY	93	-----	95
Db	6628	LVLERLSDAIRNGHEVRGIIRGTAVNQDGASNGLTAPNGPSQQRVIRQALADARLAPSEI	6687	Db	7708	DEAAEGYVDALARALRAEGTPALAVSNWMAELTDPSTTRHLRMNGLPLVMDADAALTAL	7767
QY	78	-----	77	QY	96	-----	95
Db	6688	DAVEHGTGTTLGDP\$IEAQALLNTYGPGRD\$DQPLLLGT\$VKSNI\$GHTQAAAGVAGVIKML	6747	Db	7768	AGAVADGSA\$VTVADVVRWD\$TFAPAHETRPTALPDGLPEARTALAGADRAREGATASDG	7827
QY	78	-----	77	QY	96	-----	109
Db	6748	LAMQHGTL\$PRTL\$VTSPT\$SHDWSSGAVSLUTEEDWPETGRPRRAGVS\$FVSGTNAHV	6807	Db	7828	YGRWILLEQPAARD\$DILLALVSEKAALVLGHAD\$TTLVEPDL\$PFRDL\$GFD\$SLTAVDLRNLQ	7887
QY	78	-----	77	QY	110	-----	111
Db	6808	IVEQAPIDE\$PVEPTE\$PAPAAE\$VTV\$PWIVSGR\$REALQDQVDRLTAYAAARPELSPLD	6867	Db	7888	TAATGITL\$PATLV\$FDHPN\$PAA\$LAHLRTELLGGATDTTAAPVAAPAA\$TGD\$DPIVIGMACR	7947
QY	78	-----	77	QY	112	-----	111
Db	6868	VGRSLATDRTL\$FPHRAVLLAGPDGVREARSVASRTRGRTAFL\$FSGOGAQRALMGREL\$YE	6927	Db	7948	FP\$GVNSPEDLWQLVLDEVD\$AVGDFPADRGWDL\$DALACD\$GPCRSAT\$DQ\$GFLYD\$ATD\$FDP	8007
QY	78	-----	85	QY	112	-----	111
Db	6928	RYPAFADALDTVLAQ\$FDTALD\$FSLREVL\$FAEPGT\$PEARLNETGWTQPALFAVEVALHRL	6987	Db	8008	GLF\$GVSPREAMVMD\$PQORILLEASWEALERAGID\$PGLRAG\$SGT\$TVFVGGSGDYRPPAE	8067
QY	86	-----	85	QY	112	-----	111
Db	6988	VESWGLT\$PDQVAGHSIGEIAAAHVAGV\$F\$LEDACALVAARASLMQELPRGAMLAVRATE	7047	Db	8068	Y\$GOWT\$AQ\$SASLL\$SCLRAYT\$FGI\$QFT\$VSD\$TAC\$SSSLVALH\$LAQALR\$SGECS\$TALAG\$	8127
QY	86	-----	85	QY	112	-----	115
Db	7048	EEVAPHLTERV\$VA\$VNGPSAVV\$VAGDEA\$EVA\$IAAH\$FEAQDRK\$TTRLR\$VSHAF\$H\$PLMD	7107	Db	8128	VTVMATPVGFV\$F\$SAGALS\$PDGRCR\$AF\$P\$DANGT\$G\$W\$EGV\$MLVVERLS\$DARRNGH\$RVL	8187
QY	86	-----	85	QY	116	-----	115
Db	7108	PMLTAFARVAESL\$THEPVI\$PLV\$T\$VTGVPAGAELL\$TPGYWVRHVR\$TVRFADGVRA\$LRE	7167	Db	8188	AVLR\$G\$AINQD\$GASNGLT\$AP\$GPAQ\$QORVIRQALANARLEAAD\$VDAVEAHGT\$GTL\$GDPIE	8247
QY	86	-----	85	QY	116	-----	115
Db	7168	DCVTTFVEIGPDGQLTAA\$AQ\$TLDASGDEAPATVVALQ\$RDR\$SE\$TTLLEGLATLHTQGA	7227	Db	8248	AQALLATYQD\$RERPV\$VLSGLKANIGHTQAA\$SGVGVGI\$KMWL\$M\$QHGLL\$PRS\$LYTEN\$PSS	8307
QY	86	-----	85	QY	116	-----	115
Db	7228	GP\$DWTAMFAGTGH\$RVELPTVAFQRYWPEPAPAGATGVQD\$PVDAA\$FAA\$VEREDLESL	7287	Db	8308	HVDWTAGNARLLTELTPWPE\$SERVRRA\$VSSFCASCTNAHLILEQ\$PEPVRQ\$PVEGARPAE	8367
				QY	116	-----	115

```
Db 8368 DGSPVPVLLSGRSONALRAQARKLLAHVAGHPEETVTDLARTLATTRAFFEHRAAATAAD 8427
Qy 116 ----- 115
Db 8428 RDALIAGLTAVAEGTTAAHIVEHQVERTGKHAVLPSGGGSOOLGNRDLYERFFAFAEAF 8487
Qy 116 ----- 115
Db 8488 DAVLARLDGHLGASLRDVVHMGDOETLNETGHTQPALFAFEVALYRLVESNGVTPDFVAG 8547
Qy 116 ----- 115
Db 8548 HSGVEIAAAHVAGVLSLDDACRLVAARARLMDLPRGGAMVAVEATEQEVAAELTEAVSI 8607
Qy 116 ----- 115
Db 8608 AAVNGPSTSVLAGAEDAVLAVALAERGRTRSLKVSHAFSHVLMMDPLAEFRKVAESL 8667
Qy 116 ----- 121
Db 8668 RYEAPRIAVSVNTCDVAGADELDCADYVVRHVRCVTRFADGVTALAGRGVRLLELGPD 8727
Qy 122 ----- 121
Db 8728 AVLALAEQSAPTVSIIPAQRNRDEVETLVAALGRHLVHAAGPRWDAFFDAAPGSMWLDL 8787
Qy 122 ----- 121
Db 8788 PTYAFQGRFNPDTLPTAPOGATAGEADGTSFAFMDAVAQEDFGSLESVLDVGDALSK 8847
Qy 122 ----- 124
Db 8848 VLPALLDWRRTSETQDLSWRHRIWVKRLTGAAWAHKKPLTGTVLAVVPEGLGEDPWT 8907
Qy 125 ----- 124
Db 8908 TALDAGTRVRAEVGADGREAMAATLREAAEGTRFGGVVSVLLALRETVDGVPEGVAL 8967
Qy 125 ----- 124
Db 8968 TGTLLQALGDAGIEAPLWCVTRSAVANSRDRPRRPLQAAVWGLGRVAALYPNRWGLV 9027
Qy 125 ----- 137
Db 9028 DLPERADERAAAGLAVALAGLDGEDQVAVRGSAVLARLVPAPDRGQSGAWDPTGTVLIT 9087
Qy 138 EDTGF----- 142
Db 9088 GGTGAIGAHVARRLAKGVQHLVLLSRGAEAPGATALRDELQTLGAGVTLLAACDASDRG 9147
Qy 143 ----- 142
Db 9148 QLAADVLAIPDDRPLTAVLHAAGVLDGVI DRLTPERYOKVFRAKVTSALLLDELTRDL 9207
Qy 143 ----- 142
Db 9208 LAAFVLFSSASSAVGNPCQANYAANAVALDALAEARRVQNLPATSVSWGANGGGGMADGH 9267
Qy 143 ----- 142
Db 9268 GADDAARBAAGVAMPDPLAVESLLRLVTGKEPTAVVAEVDLDRFASAFGGARPSALLREF 9327
Qy 143 ----- 142
Db 9328 PGYRETVAAPASPEGTLADRLAGMAPARRLDTVDLVRTRAAQVGLGYPIDAVGAERS 9387
Qy 143 --DLGV 146
Db 9388 FRDLGV 9393
RESULT 9
```

```
Q9NFS3
ID Q9NFS3 PRELIMINARY; PRT; 16215 AA.
AC Q9NFS3;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE D-titin.
GN SLS OR D-TITIN OR CG1915.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang Y.Q., Broadie K.S.;
RT "Characterization of Drosophila D-Titin gene.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AJ271740; CAB93524.1; -.
DR HSSP; P56276; 1TLK.
DR FlyBase; FBgn0003432; sls.
DR InterPro; IPR002106; AAcRNA_ligaseII.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IGC2; 15.
DR SMART; SM00410; IG_like; 34.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00179; AA_TRNA_LIGASE_II_1; UNKNOWN_1.
DR PROSITE; PS0002; SH3; 1.
KW Immunoglobulin domain; SH3 domain.
SQ SEQUENCE 16215 AA; 1841509 MW; 242C8765E00F7603 CRC64;

Query Match 31.2%; Score 241; DB 5; Length 16215;
Best Local Similarity 0.6%; Pred. No. 90;
Matches 90; Conservative 29; Mismatches 29; Indels 15640; Gaps 27;

Qy 1 AAG----- 4
Db 99 AAGNALFEGRLKGNPKPFVWTRKCAPLLESOKFRMSYNEATGDVSLINQIGPGEDE 158
Qy 5 ----- 4
Db 159 YTCARNQGEAICSVVIOPEGAMPALQIONLEKNIYNGYSYTSIEEPRVDTPEYR 218
Qy 5 ----- 4
Db 219 LLREVSFRAITRRSGVEODSOLSOELDRNOGPAOAPISOKPRSSKLEGSDAVFTARV 278
Qy 5 ----- 4
Db 279 GSNPKPLTWFHNGQRLVASQKYEISYSSGVATLRVKVNTARDGGHYTLAENLQCVVS 338
Qy 5 ----- 4
Db 339 SAVLAVEPAETAAYEPKVDVMAEQLEAGKALPPAFVKAFGDEITEGRMTRFDCRVGTN 398
Qy 5 ----- 4
Db 399 PYPEVFWLINGROVDDASHKILVNBESGSHLMTITNVRDLGAVOCLARKAGEAIEA 458
Qy 5 ----- 4
Db 459 QLVNLEKEQVAPQFVQRFSTMTVREGPTMTMSANAIGTPQPRITWQKQGVQISSTAERP 518
Qy 5 ----- 4
```

Db 519 VGIDGATCLEIPRVTANDAGHYQCTAQNIAGSTANRARLYVEVPEQPNYEORLNLPR 578
Qy 5 ----- 4
Db 579 PTKVIEPEIPGPEIIYLRHVERAKPHLRPGEDRVYPPQFIIPLQNVQOTEGGRVHME 638
Qy 5 ----- 4
Db 639 ARIEPVGDPTMVVWYLNPRPLAASARATSVFKFGFIALDLLSIMGHDGSEYMCRTNAS 698
Qy 5 ----- 4
Db 699 GVAESRAILSVVORPSIEQSSQNPNSLOYNQINQLEDYSRYQRTESIDEQLNQAPQIRPLR 758
Qy 5 ----- TLHLB- 9
Db 759 DLGEFEQGNVHFEAQVTPNDPSMRVWYKGLPITASRITAIFNFGVYSLNHLHLRA 818
Qy 10 ----- 9
Db 819 EDAGTYTVRAVNRIGEAIQSOSIRVHRSQV7ADLGIPEQORYIEKVEELEDYRKSQRR 878
Qy 10 ----- 9
Db 879 HVQEAEEATAPQFKTPIQNQLDLREHAHAHFEARLEPVGDSMTMRVWLKDGQPLEASSR 938
Qy 10 ----- 9
Db 939 IITYHNFGYVALTIKOLTIYDAGTYTCRAYNAMQDITVAQLTVISKNEIVSESQHPGGL 998
Qy 10 ----- 9
Db 999 QKIQHLEDSSRYGRREEETIYITQAPRFLGPKGTTKILEGQRAHFEARVEPOSIDLGLVI 1058
Qy 10 ----- LLVA- 13
Db 1059 EWHNGRSITAAARIQTYIDFGYVALDISQVRAEDAGVILVVARNKLGEAQOQATMIVET 1118
Qy 14 ----- 13
Db 1119 RSSIDTSSMHRGLYEKTONLENKPFVEQYDIBEISKSKPVFTPLSDPKPIHDGKNHIL 1178
Qy 14 ----- 13
Db 1179 ECRLEPMGDPTMRVWFPHNGRPVTVGSRFRITYYDFGFVALDIIKATAADSGETVTRATNH 1238
Qy 14 ----- 13
Db 1239 LGTAHTSACVRVIDHTDVVTETQNEQSLQEQIQLLEDSSRRHHQEEDITIMQAPQFTRGLH 1298
Qy 14 ----- 13
Db 1299 NIETIEGTNVHLECRLOPVGDPSPMRIEWFVNGKPVKTGHRFRPAYEFDYVALDLLGCYAI 1358
Qy 14 ----- 13
Db 1359 DSGVYTCOARNQLGEAVTSCSVRIIAKNLDLILETONESGLQIOYLEDSTRHRSEFVDE 1418
Qy 14 ----- 13
Db 1419 VVNIRPRFLTHPKSLTNTREGGHAHFECKIEPVTDPNLKVWFNKGRRPITVGHFRFRPIHD 1478
Qy 14 ----- 13
Db 1479 FGYVALDIVHLIAEDSGVYTCRAVNLIGSDETOVELQCRSGEQIVTVTONEAGLEQIHYL 1538
Qy 14 ----- 13
Db 1539 EDRSRYTRREIDESTKQAPVTTLSLKNVEIKENQRAHFECLLIPVSDPSMRVWYHNHL 1598
Qy 14 ----- 13
Db 1599 PLKSGSRFTETNFGFVALDIMSTLPEDAGTYTCRAYNAVGEAITSAVAVVHTKKSIYLE 1658

Qy 14 ----- 13
Db 1659 SOHETALPRLOHLEDGSKRQRISVQDEFVSQAPVFTMPVRDVRVAENQAVHFARLIPVG 1718
Qy 14 ----- 13
Db 1719 DPKLTVELWRNGQPIEASNRTTTMHDFGYVALNMKYVNPEDSGTYTCRAVNELQOAVTSA 1778
Qy 14 ----- VGP- 16
Db 1779 SLIVOSKTSIOLETOHEAAMHKIHQLEDHSRYQRRESEYTVTTAPVFTVKLIGPSNLVE 1838
Qy 17 ----- 16
Db 1839 QOSAHYECRIEYPDPNLKVWPHNGKPLSTGHRFRITYDFGFAALDILTYYAEDSGEYT 1898
Qy 17 ----- 16
Db 1899 CRVTNNLGEAINSVLNVTSRSIIHETOHEALTKIOHLEDTSRFORKTDEBOFHAEPR 1958
Qy 17 ----- D 17
Db 1959 QFGRPLRNAKYNEGAPVHLEATLIPVNDPTMKVWYCNCRPIOTGHRFKTTYDFGFVALD 2018
Qy 18 VFOAHOEDTERYL- 31
Db 2019 ILYAHAEDTGYMCKAKNAIGEAVTTCVAVNTANKTLDLTDLAORLEKIROLETYAPPP 2078
Qy 32 ----- 31
Db 2079 KPVVEEKQKPIFLTPLSNLEHLEKEGEHAHLECRVEPINDPNLKIENFCNGKOLPTGHRY 2138
Qy 32 ----- 31
Db 2139 RTHDFGVVALDILVYGEDTGYICKATNQLGEAVNTCNVLRNRRSMILDTQHPDALE 2198
Qy 32 ----- 31
Db 2199 KIQKLESKVPNARTEVGDAPIPPHFTAELRGSTEIYEGOTAFEAQVAPVHDPNLRIF 2258
Qy 32 ----- TNLNI- 36
Db 2259 YHNGKPLPSASRPHITFDGYVVSLDITHAVAEADAGEYSRAVNALQOAVSSTNLRLVIRG 2318
Qy 37 ----- GAELLR- 42
Db 2319 TIISDTQHPGLEKIRKLESTAPHORQEPETPGTRQRPVFTQPLQINIDRINEHOTAHPEA 2378
Qy 43 ----- 42
Db 2379 RLI PVGDPNLKVWYRNEKIIEDSSRITKQHDGFGFVSLDISHIRKEDEGVYMCRAVNLPG 2438
Qy 43 ----- 42
Db 2439 EAVTTASMRVWSEASIQMDTOHPDSISRIHQLEKPLAPRPTPEPERLFEKPIFTOLLTGPS 2498
Qy 43 ----- DpsL- 46
Db 2499 ELWEGTHAHFEARVVPVGDPSLKPFWFINGVELQMSRLRTTHDFGFTLDITAVVPEDA 2558
Qy 47 ----- 46
Db 2559 GYVMCRAYNAAGEAVSSTAMKVTKSNIDGOPLIPESWEAIRLKEAAMNRVPMFVDSPT 2618
Qy 47 ----- 46
Db 2619 QOAPVFTTHLSQYDKLHEGOHVLLLEAQVEFRADPNLRIEWFNKNGISLTGTSRIRSTFDFG 2678
Qy 47 ----- 46
Db 2679 LVTLSINGLRADDSAIYCTKATNOVGEAVSTSSLKIEDRHLQAESLHPDSLPRIGELEA 2738

QY 47 ----- 46
Db 2739 PKEGRPEAPEYETPVFITHLNNIECKESDNVRECNVEPARDPTMSIENFYNGQPLQA 2798
QY 47 ----- 46
Db 2799 AAKFKSIYDFGYCALDLTNSAENSGVYTKATNSKSGATTSGTLKCTGKTMFLDTQHP 2858
QY 47 ----- 46
Db 2859 OGEAGLEAVQETEELANRYTSKTKPETQYPPPVWTKPQLQAEFHLSEAQPIHLEANVEP 2918
QY 47 ----- 46
Db 2919 KEDPNLFIEWYFNGKMLNHSRFXMTSEFGFVTMDMIEVYARDQGIYTCAYNKAGEAFT 2978
QY 47 ----- 46
Db 2979 STTIFCSSKENIESTQHPKGAEGLEQIQDLEDSLKDGSKPEQDLGIPPRFTTFEVNI 3038
QY 47 ----- 46
Db 3039 ADIGEGELAHFEANLIPVGDQSMVFEWFYNGKVLASHRVRTIYAFGTVALEVLTKIED 3098
QY 47 ----- 46
Db 3099 TGYTYTCRATNKGHTAEISCNLECVDKPRQKPRFTSHIQPLEGLKDGQSAHPECTLI PVN 3158
QY 47 ----- 46
Db 3159 DPDFLVEMYHNGKLMRHSNRINKTVSDFGYVVLDISYLDQDSGEYVCRAMNKYGEDFTRT 3218
QY 47 ----- 46
Db 3219 TLNCGRGGVFYDSLQPSLORIRELECPGQQAADTAPLVAEPKFITQIVDVTKLVEG 3278
QY 47 ----- 46
Db 3279 QSAHPEARLTPTDPDLVVEWYFNGKKLPHGHRFRTFDFGIVILDLICYEENSGVYEA 3338
QY 47 ----- 46
Db 3339 RARNKYGEDVTRASLKASKSSLIILDSQLPRGMEGLEKIANLEYSMVRTREETTEETKG 3398
QY 47 ----- 51
Db 3399 KAPVFTVPLENIENREGENAHFEARITPADDPKLKVEWYMGRLKAGSRFRFTCDFGF 3458
QY 52 ----- 51
Db 3459 VILEISPVYPDSGEYSRAINNEYEAVTATMKIQKRSIIMESQLPKMEGTIDRIAE 3518
QY 52 ----- 51
Db 3519 LEGLSRSTEFVDDDTGKPPFITSPFDMVIGENALAHFECKRLQIPNDPSMRVDFHNG 3578
QY 52 ----- 51
Db 3579 KALWAGSRIKTINDFGFVILEIAGCYQDSGLYTCKATNKGAEATVSKLVQKRGQIVM 3638
QY 52 ----- 51
Db 3639 EPOLPSNFRGTGTSLOKLEETHMKREELVTEDEQPNPPKFTTEIKDNLDPPEGPIHFDG 3698
QY 52 ----- 51
Db 3699 RVEPVGPTMRIEWYFNGHVMATGSRVHQLNDFGTLVDVYIYARDSGEYTCRATNKGW 3758
QY 52 ----- 51
Db 3759 TATTSKAVTCKGHNIVYESQLPEGMTSEKLELGRIRIPEAPKVVEEVFGPPKFTTQIT 3818
QY 52 ----- 51

Db 3819 SVTVDEAEAVRFECQVBKTDPSLRVVEWYRNGKPLPSGHRYNIFDMGFVSLDILYVYGE 3878
QY 52 ----- V 52
Db 3879 DSGEYVCRAINNYGEDTRATVSCCKLPTILLQNOVPRGMRKRSALTQMEATIKKYTSEV 3938
QY 53 HL ----- 54
Db 3939 HLTEDDLDPDRKQPPRFVTQIKEQLTTEMAVTKPECOLAPVGDPNMKVWFNGKPL 3998
QY 55 ----- 54
Db 3999 HKRNFQPIYDFGYVAMNFGWYVPEDSGEYVCRATNLNKGKOBETRAIIKVSGKPGIYVSOL 4058
QY 55 ----- 54
Db 4059 PAHQSIIDRIREMEASQWVPDEVPDAKPRTKPVFVSKLEPQTVESGDPARFCVRVTGH 4118
QY 55 ----- 54
Db 4119 PRPRVMNLINCHTVVHGSRYKLTNDGMFHLDPKTRQYDTGKVEIARNSVGESIAATTEL 4178
QY 55 ----- 54
Db 4179 KVARSDDYRNVLKXSPRWYDYELAAAYQKEROENELEKVFDERKQVLSEOSSHTLKGVE 4238
QY 55 ----- 54
Db 4239 HLKPKQYKPTPDWQNVKAKKSEYNNKLTLEBOLLKETNLRRDTHQYAIPEGKVVS 4298
QY 55 ----- 54
Db 4299 SSQAKGMAQSYEENLOEKTSTTEVQAAPPKGIAQPSSESVHGREVHMNQOQVKEIQGD 4358
QY 55 ----- 54
Db 4359 LEITRKITATETTEVEHKGTTIOERWQGVKPAKAPVTRKIOPCRVFENSOAKFEVEFE 4418
QY 55 ----- 54
Db 4419 GEPNPTVKWYRESPIQNSPDLOIHTFSGKSIILIRQVVEDSAVFSCVAENRGGTAKCS 4478
QY 55 ----- 54
Db 4479 ANLVVEBERRRAGKGIQPPSFVTTIQSTTVATGOLARFADAKVTGTRPLDVYMLKXGMKI Q 4538
QY 55 ----- 54
Db 4539 PSIKFKMLEDSVHTLLIIEPFAEDSGRYECVAVNAAGEARCDGDCIVQSBKPEKPTTP 4598
QY 55 ----- 54
Db 4599 GSEKAPHIVEOLKQSVTEGSKVIFRCRVGDKPTTARMRGENFVRPSRYFQMSRQGEY 4658
QY 55 ----- 54
Db 4659 YQLVISEAFPDEGTYKCAENKLSIQTSQALKVRPIENLDAPPITALKDVSVTEGMP 4718
QY 55 ----- 54
Db 4719 AQFKTTVTGKVKATSVQWFREGQLIPETPDFQMI PDGNSAVLLICTTVEEDSGIFTVRVT 4778
QY 55 ----- 54
Db 4779 SSTGOVSESSAKLVKRRISAFQRLTIDSAEDESSESSSREDSAPESPHAFOPGQOQOF 4838
QY 55 ----- 54
Db 4839 GQFLGVNCGOHOGRSROKKPKVRSKSLQPATKVI PHRKSSRPTGRSLDKGVFLPGFKP 4898
QY 55 ----- 54

Db 4899 EPVKSWEETINLKATPIEKKKPAPKLEAAKVVLKSIKTERDQGIMSLGATLEQIAGKT 4958
Qy 55 ----- 54
Db 4959 EKEAIPWITMREKLKAVESVQOQLNKFDLDEVYLQPLEGGIETEGOLPQQAQVEQVQRTK 5018
Qy 55 ----- 54
Db 5019 EIQRKSMESVEIMEMTDQDKLITQOONAKDLIPWKEMRQOLKSQVQRTKQIDKFKIEE 5078
Qy 55 ----- 54
Db 5079 VELRHLQAOQAITEEYQGTAEETVVMIDESSKGSISKVLRDRDEQLQYEDQSNIIYKQKFI 5138
Qy 55 ----- 54
Db 5139 TTEDVINHVSEKLEAORLIREQAVNMRQOQORPQLOPLTSVEDTVISQTSERQKLV 5198
Qy 55 ----- 54
Db 5199 QOQSFISEAORQFQVEDSQMMSLEEYEHQKIINORTQOEAFSWRQPREPQKFIQVEDS 5258
Qy 55 ----- 54
Db 5259 TLLHLQERHDTQEQLLQQPVMWDRGKKPDQPPQYVQPOEQRVKKEEFVEKPKTYEEMHD 5318
Qy 55 ----- 54
Db 5319 ELVEPTIEQOPVPVMMWGRGKKPKPOEKTFEAHDDELVEPTVQOPEPVPVMMWGRGK 5378
Qy 55 ----- 54
Db 5379 KVAQOETVLSQEVVQTSQVVEQQIVBETKKTAVRRVIPPPEPEQKVEQVTLKPTPRPRPK 5438
Qy 55 ----- 54
Db 5439 EAVKAEIQLPLRSTRPVPQVEAQKAYEATDELTEPPIPOPQPMWGRGKKPKQKP 5498
Qy 55 ----- 54
Db 5499 QEEVTEIPKTLIAVDTLEEEVPKTEPQPPQVLMARGOKKQKQDEQKELPKSLIAV 5558
Qy 55 ----- 54
Db 5559 DTIEEDLIPQPEPQVPLWNERKKKPPQPDVIEEKLDAVPTKTYEKAVDVLDPKPKVEE 5618
Qy 55 ----- 54
Db 5619 KPEPVLWQGGKKIPKSEPTVEVHPDEVDAQIETVVKDEMIVEEKRRIKKTKRPKSTKE 5678
Qy 55 ----- 54
Db 5679 VTEELFEQPEEISPEEVPQKEVIEEIEEVEKKRRLKTKPKLTOQVTEETPHEE 5738
Qy 55 ----- 54
Db 5739 IIESEBEVQOEIEVEEKKVKKPKTVAEKQLKEEIEPTTEETVEEETAEADQQLVV 5798
Qy 55 ----- 54
Db 5799 EESKKVKKVKPTGCTVEKTDVEELPGBEVPVEVPVEVPDEVPABEELIEEQEIEVDQD 5858
Qy 55 ----- 54
Db 5859 EIOEQKRVKAKKPKTKTIEKTEIEIEDQPEBEVLQEEIIGEOEITERQKRVKSIKKP 5918
Qy 55 ----- 54
Db 5919 KVVTEKTVDQTEQPEKPEESQAEVKEVTYTESPKPKPAPESAQVQVEKISLKAPRK 5978
Qy 55 ----- 54
Db 5979 QRLLPKEQVEVLLKPKVKKIVAVSEAEQPETPETEPEVKEFAITTTEDILDVTKKRVKK 6038

Qy 55 ----- 54
Db 6039 KKPKTVAAEESTEPABEETEEFEEBEOEQVEVPVEIPEBPQVKEVADERKTAPKPKP 6098
Qy 55 ----- 54
Db 6099 RKBEIEIKVEEVALKRVTRPKKELPOEATIEEVRUKPTORTSIKPBEVKLEBEVDLOHVEK 6158
Qy 55 ----- 54
Db 6159 KEDEIVQEBEKTRKRVKKPKKHEDLPEIPDAEPTQLEEAHEIELEKQKPEEDQPOQPMWR 6218
Qy 55 ----- 54
Db 6219 GEKKQPEEVLBEKKWPSGKRRLPEQOPEEVQLKPIPSKPIEEQOKPEKAIPGQLVPE 6278
Qy 55 ----- 54
Db 6279 EKPESEEELELEPLKLPEDKKPKPKAKKEKKKKPKLKKATPSVDEUSEEVAEPDEPI 6338
Qy 55 ----- 62
Db 6339 ABEDEVEMPVDDVKVAVSEDLVPEEVEVPTEETPEAKQAKHKKRTKRLKEASVEGQPO 6398
Qy 63 ----- 62
Db 6399 LLEAAIAEIEKVEISQISQKTIITLLKKTEDTRPQFITTEQLIIELDVEDVRRDLEMVKT 6458
Qy 63 ----- 62
Db 6459 SNIKKERVRVLDSDSPLPELELITQKRIQEGIDKVADEELIEDOQLIQOQETTSEV 6518
Qy 63 ----- 62
Db 6519 IGOERKLVKKKKKEIKPPRITEKLPRQCVPEEPTVLECKVEGVFPPEIKWYFNILLFA 6578
Qy 63 ----- 62
Db 6579 SEKYEITVMEQVAKUKIAKVTSPDVGVYTCBAKNBAGVATSRNTIILEKEQCVPOPTKP 6638
Qy 63 ----- 68
Db 6639 LKIEFIEEKQPERLKVTVTCQVTGKPNPEVKWYRGIEEIVPSETVQMPFYDEKTDGDALEV 6698
Qy 69 ----- 68
Db 6699 INPTNEAVVYSVQAQNOGRAIGNANILSRVDEVPREILKAPTVTPLSAVVVPTGTLF 6758
Qy 69 ----- 68
Db 6759 FEAKYDGLPRPEVKWNRNGREIIEEETIETTETTTIKVNMTRKRTGKYEVWAKNKV 6818
Qy 69 ----- 68
Db 6819 GEAKSSGSVVSDQKPDQIKPPRFIQPLEPKYFGEHEVAIIIEAIVSEPLSSFQMFVHN 6878
Qy 69 ----- 68
Db 6879 EPIKSSNEVRIVSQANKSTLLIENFQSKFVGPFCTCAENVGSGVSTSTATVNLIPOEAAE 6938
Qy 69 ----- 72
Db 6939 FESPRFVEELVQPEVMQGEALLLTCQVTGKPTPKVWYHNAEKIENKETTISODLOGV 6998
Qy 73 ----- 72
Db 6999 COLQITEVPFNEGOEVCATNKIGKSVKTNVKIQAFEIPDSSEITGIGTSEEDLLORT 7058
Qy 73 ----- 72
Db 7059 LSIDEQAPKIIKKLPEKIEPKEGEQAKLEVWVGPKPKVWLRDEQIFASEEYQIENF 7118

Qy 73 -----LTSSLSV----- 80
Db 7119 EDGTSVLNVHYPDGLGISPEAYNPLGVAVTTALFAVEGIVGSKOYRKPEWSQMEEM 7178
Qy 81 ----- 80
Db 7179 QVALKAACSPSLNEMRDCRAALGETAKFSIQFAGNPIDIQWYFNNVQLRASEKYRW 7238
Qy 81 -----CGW----- 83
Db 7239 VQEQTALIMKITSEDGCGYNCKLINEIGMTMTAKFDISSTISIVSETKAKTTVKKS 7298
Qy 84 ----- 83
Db 7299 GKKTWKRSGASEQNVQKTEIRIIPTSAVETSMNVKVPVSLVEKSEISEVLVVKD 7358
Qy 84 ----- 83
Db 7359 REVADAERSOLIEIEIEEIEEKVQHDDEVEVQEQETYTSKKIEITKTVELI 7418
Qy 84 ----- 83
Db 7419 RTKISEKIITIEDVQVLSHHEEVQWLLSEIAESFGQIGESALRDLATIGLLRYGCEHY 7478
Qy 84 ----- 83
Db 7479 EITYWYEQNIFISKPKPESQALVOLVEREGHEELISQILSESSNEDETILAATVGFKPF 7538
Qy 84 ----- 83
Db 7539 IRMIQYETIIEIVIRKFVREDFISQDWKICCKERIVETSQIIESHEAITHVKIETATTK 7598
Qy 84 ----- 83
Db 7599 VEKLPKQEOEHVQNLQEOEQVKIQVQTKQIAQMTKIKKHKKHQEQEVSSETTIQCEQ 7658
Qy 84 ----- 83
Db 7659 KETLAHETSAELPQSETLEQIESLSTYETLPIQNLSKDTLQTVAVSVTTELSTPSPTAS 7718
Qy 84 ----- 83
Db 7719 RVOEILPQKVLAINEEVLPLDEFGLRKESPRKPKENKLTENIEVRLKHALNVSHAKTAES 7778
Qy 84 ----- 83
Db 7779 SKELPSKIPKSVKQAKMKESRSLVVEAPNABEEAIEDLKPKAVSQEQVQSDILFSHITE 7838
Qy 84 -----SQT----- 87
Db 7839 EQHQALETIEKLKPTSAIEDTVQOKLLSQEELIAEVLPSSETVGRDVTDRPPGETISPR 7898
Qy 88 -----NPED----- 91
Db 7899 LTPNMSLCITECPEDSIGEMQAAKERMETPSMSVTESKAVGGOEVLNVDMHPLIT 7958
Qy 92 ----- 91
Db 7959 OPTKGLADYTIKAEVPVQOEIITFDLSLRETVAKTQTAKSNALELPELSEGLVSSTAD 8018
Qy 92 ----- 91
Db 8019 SHSPIAEDLPIFEKDVKEATIDMQHQHVTTSSETVSNENAVKDKAVDTPKMAEGTLQGS 8078
Qy 92 ----- 91
Db 8079 SALTIGETQOMNLVETTVELIEPNVESTKPAKGALTEAYGTAESNEETLLESGLVPODN 8138
Qy 92 ----- 91
Db 8139 RKIEGKVNISEGEYAKVQTTVTDTDEGEFVSVAPKLVNPKFDFVQSALQIKODTVE 8198
Qy 92 ----- 91

Db 8199 KEEILSSNIELAPOLATSNMPPAELKVTSIYEVOPGLTSSDIIITEOTKSVSANQVFETMS 8258
Qy 92 ----- 91
Db 8259 IGVTSKPDMLESTSHIDAFQHPFKTGTDLDENQOQPLEVTNVQITBESSTDIIIDVLPNQK 8318
Qy 92 -----DTDP----- 95
Db 8319 LTKAETVTDGPKYAEGLVVLPMESTIDKTEDTKPTAVNADISMHQFGQFQDREOEPLEST 8378
Qy 96 ----- 95
Db 8379 LTRTEDLKPOQTTSQFGLLSLETSVCVTEGESVLSVKERHPEOSAAIGTSSALQVA 8438
Qy 96 ----- 95
Db 8439 NITRPQHMESLDRLEQKVPPYQANVNIIEITLPNVEKIDFVLSDLNTPDYNKSKGR 8498
Qy 96 ----- 95
Db 8499 VOLJESTTSLKTTTAVVSESTEELKDLNITOPVHIKPKPYESDQKISIEQTNVLEHVSS 8558
Qy 96 ----- 95
Db 8559 LNPVLPALETIQSSIKSLHEINVRETIDILEKEESLKDVHDHISGRLAKIILDCTTGIAQVR 8618
Qy 96 ----- 95
Db 8619 QEETLEHEEDLKAPLIPEKAI PASSELHRLPLTEYVQEQQTSMDTDFKVSNKCAPNI 8678
Qy 96 ----- 95
Db 8679 DHLVETKSSSEMIYVDSSINSVDSEFPAGIVPKSLVFPFRHTMTVENAFNASENFEILSA 8738
Qy 96 ----- 95
Db 8739 DQIATNVQDSLSQSI IAEQDIAFETEQNLGLETTPTTHPKLLKDDQNLHAKLVDEATVY 8798
Qy 96 ----- 95
Db 8799 EAMQEQOKVKYNIQQAIEITHDLPOVYATDLOQTFEAEKITTREQSYVAATTDIIISRL 8858
Qy 96 ----- 95
Db 8859 GLAMTTKTHPVEGIDVLLSSPPKPSLAQTNVEETQHEVVRVRETQAI SESEBELTDGRLLPV 8918
Qy 96 ----- 95
Db 8919 SAVESIDSTFKVTSDSQQPPVFDKELSIPTVSPLEARAKPSLNLOGTTTTQVIPLESSV 8978
Qy 96 ----- 95
Db 8979 LLKQTHVAKAQOEYVAQVESNKVHVQMONLVNKHEDI FENABIEFNFKPITEGTOLET 9038
Qy 96 ----- 95
Db 9039 VVIEWPIDNVGGIHLAPQPSLTLATLSTDIVNQSHVIDTQVPLEMESEQAQPLDNIQA 9098
Qy 96 ----- 95
Db 9099 ARIKSAEDHVHTNVSEDTAAADELQSLVTEEVVSVSSIOETVELKIPLOKTANLTQOT 9158
Qy 96 -----GHAD--- 99
Db 9159 PQNSVNVCOOLAYBETPDIAFEPHALTRATTSSVPTFLKPAENATVNIYENIEHGDKPK 9218
Qy 100 ----- 99
Db 9219 GTVNLTNSNLSNLSLVVSVQEVTSVPFSLGSLATVEPOELKAMPVTKSSTNLAYSEEKVG 9278
Qy 100 ----- 99

Db 9279 NQOEVTKIETVEEDDKQPETTVTVEELPFQOEKPEEIQEIPBEVRVTVTETDGPKKKK 9338
Qy 100 ----- 99
Db 9339 IRTRVKKVKGDKQEVTKIETVEEDDKQPETTVTVEEVPVEEKKPEEIQEIPBEVRVET 9398
Qy 100 ----- 99
Db 9399 VTEDGPKKKIRTRVKKVKGDKQEVTKIETVEEDDKQPETTVTVEEVPVEEKKLEEQ 9458
Qy 100 ----- 99
Db 9459 ELPBEVRVTVTETDGPKKKKIRTRVKKVKGDKQEVTKIETVEEDDKKAETTVTVEET 9518
Qy 100 ----- 99
Db 9519 ELSAPSGVKQLKRVIVQKPEDAVTFELPERKSVILSEKEDGTPTKTVIKRIKKIQ 9578
Qy 100 ----- 99
Db 9579 GPNNEVTKVQTVVEYEKAPOTIVSVEKFNTPFPPELPEERLSEVVMPLPDEVFSEAVDEG 9638
Qy 100 -----L 100
Db 9639 RLKWKTKKRIIRKPALDNTVEEVEIGIIEQDNVEPIYSVKIQERPLTESKPEDSKLIEL 9698
Qy 101 VLYITRFDLELPGD----- 114
Db 9699 PEHVTELNVILPDGKKRRTVKSRAPKSLDDDLDEVTTIHIIEEDKEPLTKVNIIEVVP 9758
Qy 115 ----- 114
Db 9759 SDEISITPIEELPEETVTEELDENKPKKTKTRTFKKGPPDDDEYFOQTIDEEG 9818
Qy 115 ----- 114
Db 9819 KEPISLRVSDENIADIIDISKLDDKVLKHQKPHKHQDQYKYVETITEPEEASADAL 9878
Qy 115 ----- 114
Db 9879 QKPTKDKTPKQKTLTPIEEVDETVIIDEGTGEQTDQIAIKRPRKVQGNVQVEAVDEK 9938
Qy 115 ----- 114
Db 9939 PIEKKEKAKKKVVKTRDEMDDYIHFILHQBIPKTVLPQVQRTMELPQARRDSSFQ 9998
Qy 115 ----- 114
Db 9999 PVKLTPMKIEKVEFKPKQWVEISSVVEFPQMLKLKAPKQRPQEBKKKKNEASFKNKKLKS 10058
Qy 115 -----NRQV----- 118
Db 10059 WIRFPVAPYCFYVVTLETNREVGLSRNVDEAEVLKLRPKFKHKSKEKAELEAD 10118
Qy 119 -----RG----- 120
Db 10119 LGAYESDHSKNEKLLHPKYKRGKKEKIETPDESRLKFKGKVPQNEAESEVNLKPV 10178
Qy 121 ----- 120
Db 10179 KLDIAEIDAEMVPTQEEVVKKKPKKSSKPEEGLQFEPIEFPEMERTSDIREESDTS 10238
Qy 121 ----- 120
Db 10239 VSDVTSTQEKPIYKKKVTSPQKNQYKILFGOPREFEETPEDDLNRKQGERPDDDK 10298
Qy 121 ----- 120
Db 10299 ADTKLKPFPYFVLDEPGVQAEVPLSETIAKEPKKRIKVKTEQEDNTIEIVPLSP 10358
Qy 121 ----- 120
Db 10359 EDNDEQIFEITVTSSEIPOGDAKAKTIGKKVKRMNKQELDDFVTELOEPEPQEVYETRM 10418

Qy 121 ----- 120
Db 10419 SDFYEVKLTLPSEMSDQPKTKRLRHEKGDVQVLEIVSVAPGEEFFYEINVISSAN 10478
Qy 121 -----VTOL----- 124
Db 10479 TEGDSEBITTDKIKKSRKIKKDDLDAYIQOLINAEIPVTELEKEYEKIDVDGKAKPKKL 10538
Qy 125 ----- 124
Db 10539 KAKTKKPIIDEGETLQGVTEHEPTKLUKTKKPEEKQNVIEKELAEHAEVPEYDEFLINK 10598
Qy 125 ----- 124
Db 10599 TESERPOEKRVETEKDIVPIVDKVLADNLCLPFVVVEEDLKDMLATDVIALEDEKII 10658
Qy 125 ----- 124
Db 10659 RKRVRKAKDSKQYEIEIETEKPGDIPDEARVIVITTEVSGDTIDGPAPSTTEAPKXSV 10718
Qy 125 ----- 124
Db 10719 RKVKKEKLUKEFIVNIVEEAPLDHVEIYEDVLRTPPRESSEKEDIPSTTTTVEDEIVNP 10778
Qy 125 ----- 124
Db 10779 VLPEKIKTVDDVRVPKDKKKKIDNQKIKISEFEPTPTSEDSTIBEYTPKLSEHDEDLQT 10838
Qy 125 ----- 124
Db 10839 DEYSVDVKDPSLPSKKKSTKKOKESLPGPISLYTIRIETTPPEPITEKIYEDGKEVVRV 10898
Qy 125 ----- 124
Db 10899 INKRRIKKAGPKPEYLIIEVIETYEDNNPEADVVRTIETTPSIDSKPOEDHKIQVVOEKK 10958
Qy 125 ----- 124
Db 10959 PKTESLDNVIQKLIDOEIPOVDHKEFKATVLETSPESKAKKIKKHHKKTTEVIDGIPIT 11018
Qy 125 ----- 124
Db 11019 VIEVTIQTETDDEDPKPEVTLKEIDHENASEAPKVLKSVSEKPKSKKESLEFKIA 11078
Qy 125 ----- 124
Db 11079 EEDKPKPVLEDISEDVQVQIIEEDGTPKQVEIKKKVSRKHGPKQEVFEITETKAIDEP 11138
Qy 125 ----- 124
Db 11139 LSEVTVEITDEQPOBEVLPAQEKPKPKKQKLKPEDVNTYVVKLEELTEPTQFETIPE 11198
Qy 125 ----- 124
Db 11199 DADDKPQPVIEDISENVQVQIIEEDGTPKQVEIKKKVSPKHGPKQEVFEITETRPSDE 11258
Qy 125 ----- 124
Db 11259 PLAETVITELTEGLNKDVIPOEKKTVKKPKKLKPEDIQSVYVIRVLEEFNEPQWAPASTE 11318
Qy 125 ----- 124
Db 11319 KPITIEDIAESIEIVPTEEDGITKEVEVKKKVSRKQGNQVFEIETKTSDEPLAEVT 11378
Qy 125 ----- 124
Db 11379 ILESGDKSOEVTILPKEKKPIKKTIKLKPEDVESYVNVVLEEFCEPQSFESPEPTSEGA 11438
Qy 125 ----- 124
Db 11439 HETKTKTKPKPIVKAPENVILIEEMAPETVNIENVEIGBEVQVKTTLKKKKEGPK 11498

Qy	125	----	124
Db	11499	EYLIEIKETYEENKPEGDIEITTTTELVEGSPDASDQPVWQIKKKPKVKDDLOKYI	11558
Qy	125	----	124
Db	11559	QOLIEQIEITKPLBEYEPETEMDSKKPKVKSHNNKTIEVIDGLPVTIHEFNVEDIVSE	11618
Qy	125	----	124
Db	11619	PEDMETPKTLLEIKELPOLPDDSSKYLWNISDEFGADKPIKOPTQDQPIKKEKPLKKK	11678
Qy	125	----	124
Db	11679	KDVEYVPSLEAFDHTVKVWSEPTLEGTVEVTVKRKVSRKSGKDHIFEITTTSEDRP	11738
Qy	125	----	124
Db	11739	TAEVTVELSSDEVLDSEKPKHERKIVKPKQLKKDDVEEYIINIIEEPIQIPVGLVE	11798
Qy	125	----	124
Db	11799	DEVEKQKEETKPKKSPITVIATEQEDNNDNYDALVKEDLDQPIERALEKPPSPLEYTI	11858
Qy	125	----	124
Db	11859	SVEEDSVGEQKQPKPKISKPKSIKQPSVDKSPDYLNVNISEESIIDEPIPEDVWTEA	11918
Qy	125	----	124
Db	11919	ABEKPSEBPTPKVEBELETEAVEKEVTDDDKGETTKQSVTRKIKKLVGPKBEEIIEIVETK	11978
Qy	125	----	124
Db	11979	TGDTPEYEVITTEEVQKSKEAPEEKKAKTVRKAKKIPKDDLDQYIOKLLIEQDIPTKEL	12038
Qy	125	----	124
Db	12039	EKYEKIDLDEPVKMKRKPIKKVKQSEEQPKBETEPEIDKPVKEKISEYSEVSDSEPKLTV	12098
Qy	125	----	124
Db	12099	AVKEPIEKPBEKPEIIVLEETVESKREPDBEGKVEKVKTKIKQNRGSEVVHDIV	12158
Qy	125	----	124
Db	12159	EEIDDTNESVITVTTPETPDQDQPSVKOKRTKKIKKDEVEDFVKRVIEEAPQEG	12218
Qy	125	----	124
Db	12219	SVDLVVIEDFVPKPSSEKRRKKPIKDKHTSVBEETPHEDEVLLIESVPDPSLSDLLTV	12278
Qy	125	----	124
Db	12279	VDSVPIEBEENKVNQIEDTKPEKKKKPKPSAKILEENVPEDTVEKPLEALHTSDLEK	12338
Qy	125	----	124
Db	12339	PVQFSFSISKEEQKHTHPENKKSSKISSSQPKOPSTEQEYISVTEHDLKPEEKPTV	12398
Qy	125	----	124
Db	12399	QVIQSETNVEETKDDTGKVKHQTTRKMLRRPAGEGEIEIIEVVRDDQPEAEITIVIEYEP	12458
Qy	125	----	124
Db	12459	EPVNODEKPKPKKTRVKKDDIHDYIOKLLIELETPKTELEKYEKIEFEPVVKOKPLDS	12518
Qy	125	----	124
Db	12519	PIDVLDESPEKQKDKKRTICKGOCVEEBAPEQPEIPVQILEVKPVEDVKEVITEDG	12578
Qy	125	----	124

Db	12579	KPVOEKTTRKVLKKIGPERQTTFKITMIESEONDSTVIVUDEEPEIASPOSIEEHPEQSK	12638
Qy	125	----	124
Db	12639	EKLAPKPKTVRKVKKDDLSYVVKLIEEIPKVDLEKYEKVEKPEKVLTVSDSIPEE	12698
Qy	125	----	124
Db	12699	PKPKSQPISVLPDPTTKPKKTKTPKPTKTEDTQQVPDEPTETTVDTDTPELTPTQTAQ	12758
Qy	125	----	124
Db	12759	PEDTATAQITPSAQEESKSTODDTKDTIOKTVKHKKTKPDTOKSVSELSPEVHKDYOISI	12818
Qy	125	----	124
Db	12819	IHSELVEEEOPEKILEVRVIDEVAEVEESQPIVEEVEDEBPQATEETVEDVTPKSKKK	12878
Qy	125	----	124
Db	12879	KVKKKTDDHDELIKKMLEOEIEKTELEKYEKIEFDVPKKLKPEFAALEPIKIERKEQKP	12938
Qy	125	----	124
Db	12939	TKVTILDATDVPKTVKLPKSKRKEKPABELTVOLPKFRLKARMLVEYPPAPLIPKTTDI	12998
Qy	125	----	124
Db	12999	GAIKNGELSRNIGEABEILKPKPKTKTKIKKIDDLKLEKYEKYEKVISSEBEBEKP	13058
Qy	125	----	124
Db	13059	YKPEKAPKEEQEDVKLKLKGKKKKPEEAPENVTLKNIPOKPOVEEVEELKOKPK	13118
Qy	125	----	124
Db	13119	EVEIVEEQTKPKDGEFVVEFPFSEFDRPEYVPDELEQIEHPEIPEKVKPKSKTKYKPK	13178
Qy	125	----	124
Db	13179	DKSKSEPTIVSEIVAGVPKEEBAIPEQDVKFRKPERDAPETDSEIKLRVPQASKDEN	13238
Qy	125	----	124
Db	13239	PDEQALVTPKAEPIPOEIEDKAIDDEKPKKSKPKKQPKQEOBIAKEEPEEPEVSVKEE	13298
Qy	125	----	124
Db	13299	EALVDKPIEIEKPKDKVKKEKKKPEAPVSEVVVIEBEPKPEEVEEIPVEKITTITVLEP	13358
Qy	125	----	124
Db	13359	EDAPKEHQVKVIDFDERQETTEVIEBKVTRKKKPKQPOPEBFEVTLKEPKEBQIOPDV	13418
Qy	125	----	124
Db	13419	VSAEISLPIEBEPEKQEQVEVELKITOTTPBEPNDVOIAVKEKVKTRPKVKKEDKIVV	13478
Qy	125	----	124
Db	13479	EAEEEKQPVBEETIVEVEKQBEKKKSEKPKSYEFKISQTSIEEKPIEVAEBAPEETPKV	13538
Qy	125	----	124
Db	13539	EKKVAEKFSYEFTLKETDBEKVITVDDQPEEAPVVEVFKKKKPEAVEAEAFVMTBPK	13598
Qy	125	----	124
Db	13599	IVEETSVETAIKQKTKKKDBEEAQLAIKWVESEAPVABEVFSEAPESKIVEEIVAE	13658
Qy	125	----	124


```
QY 145 ----GVTI 148
DB 15879 YRSVGVTL 15886
|||:
RESULT 10
O8XS40 PRELIMINARY; PRT: 6889 AA.
AC O8XS40;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-JUN-2002 (TREMBLrel. 20, Last sequence update)
DE Probable peptidase synthetase protein.
GN RSP0641 OR RS05860.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Plasmid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]_SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502 (2002).
DR EMBL: AL646079; CAD17792.1; -
DR InterPro: IPR001227; Ac transferase.
DR InterPro: IPR000954; Aminotran_3.
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR002103; Bac luciferase.
DR InterPro: IPR001242; Condensatn.
DR InterPro: IPR000276; GPCR Rhodpsn.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR003880; Ppanctn_attach.
DR Pfam: PF00698; Acyl_transf; 1.
DR Pfam: PF0202; aminotran_3; 1.
DR Pfam: PF00501; AMP-binding; 5.
DR Pfam: PF00296; bac luciferase; 1.
DR Pfam: PF00668; Condensation; 4.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF02801; ketoacyl-synt_C; 1.
DR Pfam: PF00550; pp-binding; 6.
DR PRINTS: PR00154; AMPBINDING.
DR PROSITE: PS00075; ACP DOMAIN; 6.
DR PROSITE: PS00455; AMP_BINDING; 5.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; UNKNOWN_1.
DR PROSITE: PS00012; PHOSPHOPANTHETINE; UNKNOWN_4.
KW Plasmid; Complete proteome.
SQ SEQUENCE 6889 AA; 744917 MW; A7F31DCFFDED1D37 CRC64;

Query Match 30.9%; Score 239; DB 16; Length 6889;
Best Local Similarity 1.5%; Pred. No. 8.4;
Matches 91; Conservative 27; Mismatches 29; Indels 6063; Gaps 27;

QY 1 AAGGILHLELLVAV----- 14
DB 328 AAYGMAEATULIAFGWAIRFOGPRCLPFSRLALQKMAVAADDEADRTALASHGSALTG 387
|||: |||:
QY 15 -----GPDVFQ-----AHQEDTERY--- 29
DB 388 HQLAIVDPETLRRCPCRGVEIIVSGFSVAQGYWRDDEDSRRSFAGEMAEPADGERYLRT 447
|||: |||:
QY 30 -----VLTNLI----- 36
DB 448 GDLGFLHAGELYICGRKLDLIILNGLNIYPQDVELAAAFESHARLRENGTIAFAVDRDDTE 507
|||: |||:
QY 37 ----- 36
DB 508 QLVIVQLEFRQPVPGMFECMASAVSMNVGVTPDVIIVKAGAIPTRTSGKIRROQCRA 567
|||:
QY 37 ----- 36
DB 568 DFLADRLPAMARWDRPVLAASPVVAPGAPVLAARSPVAVSAVVEARLAEALRGL 627
QY 37 ----- 36
DB 628 DAEAIHQEPFAYFGLDMSAVQLAEALSVMRVTVPVFWFHEHPNIARLAFGLAELSA 687
QY 37 ----- 36
DB 688 GQGRHPAEAAPAPQAADAPIAIIGLCRFPGADDADAFWDALDGOVDAIGAVPOARR 747
QY 37 ----- 36
DB 748 AAGTFDEPRAELPSQVRLGGFLDRVDAFAFFSISPLEAARMDPQORLAEVAWQALE 807
QY 37 -----GAEI-----LRD 43
|||: |||:
DB 808 AGIAASGLAGSTTGTFIGISTHYENLQDRAGSELVSVSATGNAGSIVANRLSYCLDLRG 867
QY 44 PSL----- 46
|||:
DB 868 PSLAIDTACSSSLVAVHAACOSLRDGESTLALAGGVNLVLTLSSEPPARAGMILSPDGR 927
QY 47 ----- 46
DB 928 KAFDASANGYVRGEGCGVVVLKRLSDALRDGDPVRAVLGSVMQDGRNGLIAPNGSAQ 987
QY 47 ----- 46
DB 988 AAVYQALARARLRPEQIGYIEAHGTGTALGDPTELNALKSVFAHAPEAGRCVGSVKTN 1047
QY 47 ----- 46
DB 1048 ICHLEAAAGVASLIKVLALHEHTLPANLHYREPNNPHCALDGSALSIVRSOPMPGNAGR 1107
QY 47 -----GAQFRVHLV----- 55
|||: |||:
DB 1108 RHAGISSFGGTLAHMIVAOAPPAATAHGAQWPHLLLSLSAKTAPALEAMTEAVAGOLR 1167
QY 56 -----KMWI----- 59
|||:
DB 1168 DLPDAALADAAYTHOAGRNAPAWRRMLVARDREDAAEALRARPRRVFTAQVRPAVPAPV 1227
QY 60 ----- 59
DB 1228 VMFPGQAQVGMARELYQEIAPRAVYVDRCAQVLRERAGFDLIQSLYGDGDPESHLA 1287
QY 60 ----- 59
DB 1288 LTRTEAAQPALFVIEYALARLWMQWGMKPAALIGHVSVEYVAACLAGVFSLLDALLVLA 1347
QY 60 ----- 59
DB 1348 RCRLMQSLPAGNMLSVLDAESLRAQDGLTALAANVGRQRCVAGETTAAALERRLDA 1407
QY 60 -----LTEP----- 63
|||: |||:
DB 1408 QGISHRRLATSHAFSHMTTEPILPAFADSVARLTURAPAIIPVSNLTGAMIEPAQATDPG 1467
QY 64 -----EGAPN----- 68
|||: |||:
DB 1468 YWAHLRGTVRFADGLQTLMQGPHAWIEVPGQTLASALRCPAGDALVLPGLGGEGPK 1527
QY 69 ----- 68
DB 1528 SDAVLVLSAGOLWLAGQALDWRALYAGQORRTLRAPVYPPFORERHMLDAPAAPQOADA 1587
QY 69 -----ITANLTSSL----- 77
```

Db	1588	SVEPSRAVTSMPVAAAMSSSVNSASRAQAIQOALRGVAKMLHADVARVDPDLPLELGL	1647
Qy	78	-----	77
Db	1648	ADSLMMVQAIGSTEQTYGTVTVRQLFEELTTIAADYVDRQMPREARLEAAEAPATV	1707
Qy	78	-----	77
Db	1708	PAAAVMPMPAPAPAAPFAIAVDARPDAPVPLAAIPALPOSSLERLLSQOLDALSOLTAR	1767
Qy	78	-----	77
Db	1768	QLALFOGGTVAATAGGPASVAGETSPVPTVMPAAVPPVAVAPAPATAAPGGTHKPPVPYQ	1827
Qy	78	-----	77
Db	1828	PVRTAKADPFRDLSARQRAYLDGFVARYTERTRGSKALVQRYRPVLADNRVSAGFRFST	1887
Qy	78	-----	77
Db	1888	KEMLYPVVSRSEGAFLWDPDGNVYIDLTMFGVNLFGHRPAFVQEAALDAQLKTGLELGP	1947
Qy	78	-----LSVC-----GWS--84	
Db	1948	QTRLAGEVAELVTRLTGLERVAFCNSGTEAIMLALRLARTVTRDKIVVFAGSYHGSDD	2007
Qy	85	-----	84
Db	2008	TLMVADGAGTVAMAPGLQPCASAHITVLDYGAPESLARIHAHAELAAVLVEPVQSRPH	2067
Qy	85	-----	84
Db	2068	WQPREFLHALRELTREHGIALIFDEIIMGFRLHPGGAQAWYGEADWATYKVGIGGMPA	2127
Qy	85	-----	84
Db	2128	GMVAGSAAYLDAVDGGWRVGDASYPOADTTFYAGTFCCKHPLMLVAARAVLQRLAQGPA	2187
Qy	85	-----OTIN-----88	
Db	2188	LOETLNARTAEVLRLNGVFADARVPVRVAVHCGSSFRLEASPSIDLLYYHLLAGGLYIW	2247
Qy	89	-----	88
Db	2248	EGRGMFLSTAHDADVDVRVVEIFADSVRAMLDGFFEDGAPTTPSGGGRSPFAARVGG	2307
Qy	89	-----	88
Db	2308	PASTDAPVAPALAEAAVARLSAAPPVPPATGGIRFGISFFGHYASGYDAQKYRLLFEAR	2367
Qy	89	-----	88
Db	2368	YADAGFSLLWPERHFAFGLSPNPSVLSAALARETSHIQLRAGSVVPLHHPVRVAE	2427
Qy	89	-----	88
Db	2428	EWSMVNDLSOGRVGIACASGHHPNDFVPAEAFSGSHRELMPQRIEQIQLWRGEPLVRD	2487
Qy	89	-----	88
Db	2488	GSSREIEVKLFPMPROPELPITWITVGNPDTYRRAGEIGAGILTNLMQTVAEALCNLAL	2547
Qy	89	-----	88
Db	2548	YQALVEHGHGVERSRSVLLHTFVCDEAAEARAVARAPFIHYLRSSVGLFQNMVDSLGL	2607
Qy	89	-----PEDD-----92	
Db	2608	QADVSTLSEDDRDYLLSVAYERYEHSALIGSPATCRALVERLOAIGVDEIGCFIDFGVD	2667
Qy	93	-----	92
Db	2668	PDTVLRDLQDALLKQSFETAADDAAERYPLVPAQKGIWFECQISHEAALSNTTNTVL	2727
Qy	93	-----	92
Db	2728	GLRGALDHAAALARALQOVVDRHAALRSVVEADGEHQVLPVAVEVGLPLVDFSRADRDAA	2787
Qy	93	-----TDPGHADLV-----101	
Db	2788	IGQWFDVNNHRPMDPGHGPLVRACLLRKGEAEHLAITFHHVIIDGYSQEIVLOELAAACY	2847
Qy	102	-----	101
Db	2848	RAACRGAGPLPAASPFEQVERHQAYLRSRYODRAYWRQFGSLPPSLELSGRHARP	2907
Qy	102	-----	101
Db	2908	SAPSHRARRHLLTIDGERYARLQQLSRKLGGLPMTMLAGVAVLLQRLSGOQMVIGVPM	2967
Qy	102	-----	101
Db	2968	VVGRAEGSEASLVGCTLNLVPVRCDSGSPFAEFLGRIKRSVLEAHADYPFGHLLRD	3027
Qy	102	-----	101
Db	3028	LDLRASORRPLAPVLNLRSLALPQFDALQAWLEQSPISFPDDLTTIDVMQLPDLQVM	3087
Qy	102	-----	101
Db	3088	FOYQEVLFEEHAEIERMAAQFVQLLDGIVADPACSTIGRLALLSABERROIVDMVNAGEPVP	3147
Qy	102	-----	101
Db	3148	EPSETLHAVFEAQVRRTPETAIVEHEGWRVSYAELNARANRVAHALIGLVGPDARVGLC	3207
Qy	102	-----	101
Db	3208	AERSVELVIGLLGILKAGGVVPLDPSYPRDLAYMLEDSPAVAVLAOSNTREPLGALS	3267
Qy	102	-----	101
Db	3268	PVLDLENPLEGEAEHDPQVTGLEPNHLAYVIYTSGSTGPKGVLEHQRVARLFTSTOPW	3327
Qy	102	-----	101
Db	3328	FGGAEDVWTLPHSFADFVSMELFGALLHGGRLVVVPKLTARSPQAFYALLCEAGVTVL	3387
Qy	102	-----	101
Db	3388	NOTPSAFQOLMAAQOEAPAAHRHLRQVILGGEALVGLRPMWYERAEHAGTOLANNYGIT	3447
Qy	102	-----LYI-----104	
Db	3448	ETTVHVSYPALAEADAQGTGSPIGRRIPDLRVVVLDAHGEPPVPGVTGEMYGAGVARG	3507
Qy	105	-----	104
Db	3508	YLNRPDLTAERFVWNPFGHEGRMYRTGDLGRWLPDGSLEYQGRADAQVKLRGFRIELG	3567
Qy	105	-----	104
Db	3568	EIEASLSQCAGVREAVTVTVREDVFGQRLVAYVVSGEAIEAQAALRELOGSLPAYMVPA	3627
Qy	105	-----	104
Db	3628	YVRLAHLPLTNSGKLDKGLPAPEGHAYASTAYEAPQGEVEQTLAGIWOTLLGVERVGRH	3687
Qy	105	-----	104
Db	3688	DDFFALGCHSLQAVRLVTVRVLGAEGLTALFAQPSLSAVAQAIVRGGSALQAITAA	3747
Qy	105	-----	104
Db	3748	DRSEALPLSFAQORLWLLAQMEGGSAYHI PVGLRLKGLDELDALGRALDRIVARHEALR	3807

Qy	105	-----TRFDLE-----	110
Db	3808	TRFEVREGAIFORVASADVGFALDRVDLQCGADREQTAAALSEREANTPFDEGQLIRG	3867
Qy	111	-----	110
Db	3868	CLVKLGEQEHVLLITMHHIVSDGWSQGVLAARELGALYEAIRSGGDDPLPALPIQVADYAV	3927
Qy	111	-----	110
Db	3928	WORRWEGGELQOGAYWEQALAGAPTLLSLPTDRAPAOQDYAGGSVEVVFDETLSDL	3987
Qy	111	-----	110
Db	3988	KRLSORHGTTLFMTVLACHSALLSRLSQGEVVVGVSPVANRTRSEVEGLIGFFVNTLALR	4047
Qy	111	-----	110
Db	4048	VNVGSATVSELLDRVKAKVLEAQAHQDLPFEQVVERVKPVRSLSHSPIFOAVFSWHNTEA	4107
Qy	111	-----	110
Db	4108	VDLSRLALSLEARENATAKLDIQLELAEDGRIVGTINVATALFERSTAQRVADYLQR	4167
Qy	111	-----	110
Db	4168	MLQAMVADDGQOVGRIALLGAEABRAQVLOAMNATERANWPAATLPALFEAQVWRTPDVAL	4227
Qy	111	-----	110
Db	4228	KHADOQSVSYRELDARANRLAHHLRELGVADVLVGLCVDRSIEMI VGLLILKAGGAYVP	4287
Qy	111	-----	110
Db	4288	LDPDYPQARLAYIFODAMLSVLVSKRALAORLPIANTQVVELDDAEPAWADYPPPPQVQ	4347
Qy	111	-----	110
Db	4348	GEPPQLAYIYITSGTGQPKGVAVTHQVASLVYSOCERFGVSSQSRVLQFASISFDAAV	4407
Qy	111	-----	110
Db	4408	SEIGMCLLCGACVLVAPALQALMGAALTHLDRERITHVTLPVAVLALMPEQALPADCHL	4467
Qy	111	-----	110
Db	4468	IVAGEACPPSLVRLMSEGRMTINAYGPTTEATVCATMSRALTADAPSIGGPIGNVRVYL	4527
Qy	111	-----LPD	113
Db	4528	DAYLQVPVGVVTGELYIAGSLARGYWQAGLTAERFMANPFALGERMYRTGDLGRWLPD	4587
Qy	114	GN-----	115
Db	4588	GSLEYQGRADAQVKLGRFRIELGEIEARLSQCAGVSEAVTVREDAPGEORLVAYVVSGE	4647
Qy	116	-----	115
Db	4648	AIEAOTLREQQSRSLPEYMPVPAAYVRLHPLTPNGKLDKRGKLPAPGQAYASTAYEAQ	4707
Qy	116	-----RQVR-----GVTQL-----	124
Db	4708	GEVEQTLAGIWTLLGVERVGRHDDFFALGGHSLQAVRLVAQVRTQGLGELGTTELFAQP	4767
Qy	125	-----	124
Db	4768	SLSAVAQAIVRGQGSALPAITVADRSGLPLSFSQORLMFLAQMEGGSEAYHIPVGLRLK	4827
Qy	125	-----	124
Db	4828	GELDEDALRRSLDRIVARHEALRTRFVTEEGQAVQVRVASADVGFDLDCVDLQCADREOA	4887
Qy	125	-----	124
Db	4888	LATLSEREANTPFDLAHGPLIRGLVKLGEQEHVLLITMHHIVSDGWSQGVLAARELGSLY	4947
Qy	125	-----	124
Db	4948	EAYRAGNADPLPALPIQYADYAVWQRWLEGGELQROGTWEOALAGAPTLLSLPTDRAR	5007
Qy	125	-----	124
Db	5008	PPQDYAGGSVEVVFDAELSTGLRTLSQRHGTTLFMTVLAGWSALLSRLSQGEVVVGVSP	5067
Qy	125	-----CGA-----	127
Db	5068	VANRTRSEVEGLIGFFVNTLALRVEVGATVSELLERVKAKVLEAQAHQDLPFEQVVERV	5127
Qy	128	-----	127
Db	5128	RPVRSLSHSPVFOAALSNLWNTAEVGLSLELEGLTIEGVDAQAAKFPDLELRETSEGL	5187
Qy	128	-----	127
Db	5188	AGSLDYATALFDRATIERYLGLHRLKAMAADDSQEVNRIALLDEGERTOLLESWNETK	5247
Qy	128	-----	127
Db	5248	APYPRASTIHGLFEAQVVRTPEAIAVVHEGQVSYAELNARANRVAHALRRLGVGPDAV	5307
Qy	128	-----	127
Db	5308	GLCAERSVELVGLLGLTKAGGGVPLDPSYPQDRLTMYLSDSAPVAVLTOSNTREQLGA	5367
Qy	128	-----	127
Db	5368	LSVPVLDLGGLEBAEHDPOVTGLEPHHLAYIYITSGTGRPKGMNEHGRVGNRLMAQ	5427
Qy	128	-----	127
Db	5428	QTYRLDASDRVLQKTPFGFDVSWWELFWPLLAGARLVNARPEGHDPAYLAATIEQAGIT	5487
Qy	128	-----CS-----	129
Db	5488	TLHFVPSMLQLFLDOVEAGRCQGLRRMLCSGEALSHALQORSLARFPHSELHNLGYPTA	5547
Qy	130	-----	129
Db	5548	AIDVTAMRCNAEIHPPGVVPIGRPIANTQMYVLDGHQOPVPLGVTGEIYIGGIVARGYLN	5607
Qy	130	-----	129
Db	5608	RPELTAERFVYNPFHGEGRMYRTGDLGRWLPDGLSLAYOGRADAQVKLGRFRIELGEIE	5667
Qy	130	-----	129
Db	5668	ARLSQCAGVSEAVNAMEDVPGEORLVAYHVSDAEIAEATLREQLQASLPEYMPVPAAYVR	5727
Qy	130	-----	129
Db	5728	LEHLPLTPNGKLDKRGKLPAPDGOAYASAAEPAOGEVEBQTLAGIWTLLGVERVGRHDDF	5787
Qy	130	-----	129
Db	5788	FALGGHSLQAVRLVAQVRTQGLGELGTTELFAQPSLSAVAQAIVRGQGTALPAITVADRG	5847
Qy	130	-----	129
Db	5848	EALPLSFAQORLWFLAQMEGGSEAYHIPVGLRLKGEDELDRRLSLDRIVARHEALRTRF	5907
Qy	130	-----	129
Db	5908	EVQEGQAVQVRVASADVGLTLDWADLSAEAASEHQLGLLAEAWARAPFDLEQGLIRGLV	5967
Qy	130	-----	129

```
Db 5968 KLGEQEHVLLITMHHIVSDGWSQGVLAELGALYEAYRAGNSDPLPALPIQVADYAVWOR 6027
Qy 130 -----PT----- 131
Ra MEDLINE=96186896; PubMed=8635756;
Db 6028 RWLEGELORQGTWYEQALAGAPTLISVPTDRARPPQDYAGGSVEVFDETLISAGLRKL 6087
Qy 132 -----WSCUIT----- 137
Ra "Organization of the biosynthetic gene cluster for rapamycin in
Db 6088 SORHGTTLFMTVLGAGNSALLSRLSGQEEVVVGVSPVANRTRSEVEGLIFPVNTALARVEV 6147
Qy 138 ----- 137
Ra Streptomyces hygrosopicus: analysis of the enzymatic domains in the
Db 6148 GSATVELLGRVKSRLVLEQAQHODLPFEQVVERVPVRSLSHPVFOAALSWLNTAMVYL 6207
Qy 138 ----- 137
Ra modular polyketide synthase."
Db 6208 KLDGLTIEHLDSEYAKFDLSLALREHDIAGSLDYATALFDRETIERYLGLHRLLA 6267
Qy 138 ----- 137
Ra Gene 169:1-7(1996).
Db 6268 AMVENDSQVSRICLLDEGERAQLLESWNETKAAYPDASTIHGLFEAQRRTPEATAVEH 6327
Qy 138 ----- 137
Ra EMBL; X86780; CAA60459.1; -.
Db 6328 EGQVSYAELNARNRVAHALRLGVGPDPARVGLCAERSVELVUGLLILKAGGVYPLD 6387
Qy 138 ----- 137
Ra HSSP; P25715; 1MLA.
Db 6388 PSYPQDLAYWLEDSAPVAVLAQSNTEQLGALSVPVLDLDRPLLEAEHDPQVTGLEPH 6447
Qy 138 -----EDTG-----FDL----- 144
Ra InterPro; IPR001227; AC transferase.
Db 6448 HLAYVIYTGSGTRPKGVWVEHGRSLNLDWYLEDGLRSDDTVLLVSSYSFDLTOKNIL 6507
Qy 145 -----GVT 147
Ra InterPro; IPR002085; Adh_zn_family.
Db 6508 GPLLAGRLHLGFPVDPDALVAQIRREGIT 6537
Qy 145 ----- 147
Ra InterPro; IPR004410; FabD.
Ra InterPro; IPR001899; Gram_pos_anchor.
Ra InterPro; IPR000794; Ketoacyl-synt.
Ra InterPro; IPR003880; Ppantne_attach.
Ra InterPro; IPR002364; OOR_zeta_crystal.
Ra InterPro; IPR000634; S/T_dehydrase.
Ra Pfam; PF00698; AgyL_transf; 6.
Ra Pfam; PF00107; adh_zinc; 1.
Ra Pfam; PF00109; ketoacyl-synt; 6.
Ra Pfam; PF02801; ketoacyl-synt_C; 6.
Ra Pfam; PF00550; pp-binding; 6.
Ra TIGRFAMs; TIGR00128; fabD; 6.
Ra PROSITE; PS00075; ACP_DOMAIN; 6.
Ra PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
Ra PROSITE; PS00165; DEHYDRATASE_SER_THR; UNKNOWN_4.
Ra PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_4.
Ra PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
Ra PROSITE; PS01162; OOR_ZETA_CRYSTAL; 1.
Ra Phosphopantetheine; Transferase.
Ra KW
Ra SQ
SEQUENCE 10223 AA; 1072168 MW; 8631C61BDD4E16BC CRC64;
Query Match 30.8%; Score 238; DB 2; Length 10223;
Best Local Similarity 1.0%; Pred.No.31;
Matches 104; Conservative 17; Mismatches 25; Indels 9871; Gaps 32;
Qy 1 AAG----- 3
Db 86 AAGKSYCVGGFLDSAGGFDSFFGISPREALAMPQORLVLEASWEAFERAGIEPGSLR 145
Qy 4 ----- 3
Db 146 GSDTGVMFCAYPGGYGVGNADLGFGNTAGAVSLGRVSYFFGLEGPVAVTVDTCSSSLV 205
Qy 4 ----- 3
Db 206 ALHQAAYALRQGECSLALVGGVTVMATPQSFVEFSRQRGLASDGRCKAFMSADGTGWAE 265
Qy 4 --GILHLELL-----VAV-----GP----- 16
Db 266 GAGVLLVERLSDAQRKGHQLAVVRGSAVNQDGASNGLSAPNGPQORVIRAAALSNAGLS 325
Qy 17 ----DVFOAH----- 22
Db 326 TAEVDVVEAHGCTTTLGDPFIEAQAALLTYQDREOPLLLGSKVSNLGHTOAAAGVSGVIK 385
Qy 23 ----- 22
Db 386 MVNALQHLVLPRTLHVDEPSRHRVDMTDGAVALVTENQHPMDGRPRRAGVSSFGISGTNA 445
Qy 23 ----- 22
Db 446 HVILESAPPTQAVDDVPPEAPVVASLPLVISARTLPALVEYEDRLRAYLAASPGVDV 505
Qy 23 ----- 22
Db 506 RGVASTLAVTRSVFEHRAVLGDDTGTGTTVSDPRVVFVPPGQSGORAGNGEELAAFPV 565
Qy 23 ----- 22
Ra "The biosynthetic gene cluster for the polyketide immunosuppressant
Ra rapamycin."
Ra Proc. Natl. Acad. Sci. U.S.A. 92:7839-7843(1995).
Ra [2]
Ra SEQUENCE FROM N.A.
Ra STRAIN=NRRL 5491;
Ra MEDLINE=96186895; PubMed=8635730;
Ra Molnar I., Aparicio J.F., Haydock S.F., Khaw L.E., Schwecke T.,
Ra Koenig A., Staunton J., Leadlay P.F.;
Ra "Organisation of the biosynthetic gene cluster for rapamycin in
Ra Streptomyces hygrosopicus: analysis of genes flanking the polyketide
Ra synthase."
Ra Gene 169:1-7(1996).
Ra PL
```

```
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 5491;
RX MEDLINE=96186896; PubMed=8635756;
RA Aparicio J.F., Molnar I., Schwecke T., Koenig A., Haydock S.F.,
RA Khaw L.E., Staunton J., Leadlay P.F., Staunton J., Leadlay P.F.;
RA "Organization of the biosynthetic gene cluster for rapamycin in
RT Streptomyces hygrosopicus: analysis of the enzymatic domains in the
RT modular polyketide synthase."
RL Gene 169:1-7(1996).
DR EMBL; X86780; CAA60459.1; -.
DR HSSP; P25715; 1MLA.
DR InterPro; IPR001227; AC transferase.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR004410; FabD.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR002364; OOR_zeta_crystal.
DR InterPro; IPR000634; S/T_dehydrase.
DR Pfam; PF00698; AgyL_transf; 6.
DR Pfam; PF00107; adh_zinc; 1.
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt_C; 6.
DR Pfam; PF00550; pp-binding; 6.
DR TIGRFAMs; TIGR00128; fabD; 6.
DR PROSITE; PS00075; ACP_DOMAIN; 6.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; UNKNOWN_4.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_4.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
DR PROSITE; PS01162; OOR_ZETA_CRYSTAL; 1.
DR Phosphopantetheine; Transferase.
DR KW
DR SQ
SEQUENCE 10223 AA; 1072168 MW; 8631C61BDD4E16BC CRC64;
Query Match 30.8%; Score 238; DB 2; Length 10223;
Best Local Similarity 1.0%; Pred.No.31;
Matches 104; Conservative 17; Mismatches 25; Indels 9871; Gaps 32;
Qy 1 AAG----- 3
Db 86 AAGKSYCVGGFLDSAGGFDSFFGISPREALAMPQORLVLEASWEAFERAGIEPGSLR 145
Qy 4 ----- 3
Db 146 GSDTGVMFCAYPGGYGVGNADLGFGNTAGAVSLGRVSYFFGLEGPVAVTVDTCSSSLV 205
Qy 4 ----- 3
Db 206 ALHQAAYALRQGECSLALVGGVTVMATPQSFVEFSRQRGLASDGRCKAFMSADGTGWAE 265
Qy 4 --GILHLELL-----VAV-----GP----- 16
Db 266 GAGVLLVERLSDAQRKGHQLAVVRGSAVNQDGASNGLSAPNGPQORVIRAAALSNAGLS 325
Qy 17 ----DVFOAH----- 22
Db 326 TAEVDVVEAHGCTTTLGDPFIEAQAALLTYQDREOPLLLGSKVSNLGHTOAAAGVSGVIK 385
Qy 23 ----- 22
Db 386 MVNALQHLVLPRTLHVDEPSRHRVDMTDGAVALVTENQHPMDGRPRRAGVSSFGISGTNA 445
Qy 23 ----- 22
Db 446 HVILESAPPTQAVDDVPPEAPVVASLPLVISARTLPALVEYEDRLRAYLAASPGVDV 505
Qy 23 ----- 22
Db 506 RGVASTLAVTRSVFEHRAVLGDDTGTGTTVSDPRVVFVPPGQSGORAGNGEELAAFPV 565
Qy 23 ----- 22
Ra "The biosynthetic gene cluster for the polyketide immunosuppressant
Ra rapamycin."
Ra Proc. Natl. Acad. Sci. U.S.A. 92:7839-7843(1995).
Ra [2]
Ra SEQUENCE FROM N.A.
Ra STRAIN=NRRL 5491;
Ra MEDLINE=96186895; PubMed=8635730;
Ra Molnar I., Aparicio J.F., Haydock S.F., Khaw L.E., Schwecke T.,
Ra Koenig A., Staunton J., Leadlay P.F.;
Ra "Organisation of the biosynthetic gene cluster for rapamycin in
Ra Streptomyces hygrosopicus: analysis of genes flanking the polyketide
Ra synthase."
Ra Gene 169:1-7(1996).
Ra PL
```

Db 566 FARIHQVMGLDVPDLEVNETHYAQPALFALQVALFGLLESWGVPRDPAVVGHSVGLAA 625
Qy 23 -----QEDTERVVL----- 31
Db 626 GYVSGLSLEDACTLVLSARARLMQALPPGGVMVAVPVSEDEARAVLGEVIEIAAVNGPSS 685
Qy 32 ----- 31
Db 686 VVLSGDETAVLQAAAALCKSTRLATSHAFHSARMEPMLEEFRTVAERLTYQTPLAMAAG 745
Qy 32 ----- 31
Db 746 DRVTTAEYVVRQVDTVRFGEQVASYEDAVFIELGADRSLARLVGDVAMLHTDHEAQA 805
Qy 32 ----- 31
Db 806 SALAHLVYNGVTVDWNTALLGDAPATRVLDLPTYAFQHORYWLEGADRAAAGGHPLLGP 865
Qy 32 ----- 31
Db 866 GLAEASGVLTFSQVSRSGDLWLQDOTVLPATAFAEWALAAADETCGLVEELTVEALLLL 925
Qy 32 ----- 31
Db 926 PDGAVQVQTVWSEPPDAGRRQLSIHARYSDDELWTRLATATLATTGCVSQCAGEAWP 985
Qy 32 ----- 31
Db 986 PAGAVPEARVPSLRGVWRGSEVFAEVALDDTHDTRFLHPGLMTAAFAVGEETPAA 1045
Qy 32 ----- 31
Db 1046 WOGLTLHARNAELVRVLTQSODETSLVEAVDSTGLPVLTAESLALDAPVNEPATSD 1105
Qy 32 -----TNLNIGA----- 38
Db 1106 LUTLTWAGIATRQQOTGLTVGAFEDLAADGAPVDPVVVFTALPGNDPDLAQTRILTAQV 1165
Qy 39 ----- 38
Db 1166 LRTQEWIGGERFSDTLVVRTGTGLAAAVSGLMRSAQSEHPGRFILVESDDALTPDQ 1225
Qy 39 ----- 38
Db 1226 LAATVGLDEPLRLISDGRYEVPRLTRAHADPEPERAWDPDGTVLITGSGVLGIVARH 1285
Qy 39 ----- 38
Db 1286 LVAERGVRHLLLSRGTDPDRALLSELAEELGAAVDTAACDVSDRAELARVLARVSPEHPLT 1345
Qy 39 ----- 38
Db 1346 AVIHTAGVVDDGWVESLSAQRLTETFRPKADGANHLHELTRDADLAAFMVSSAAGVMAG 1405
Qy 39 ----- 38
Db 1406 AGCGNYAANAFLDALAEERRAEGLPALAVANGLWEDASGLTAQLTDDTRDIRRGGLRA 1465
Qy 39 ----- 38
Db 1466 ISAEHGMRLFDNASRHSPEVPLVAAPMEPVDAEVPALLRSLHRPNVRRALAGGAQWLAA 1525
Qy 39 ----- 38
Db 1526 LAPEERAKLLKVVRDTAATVLGHADARTIPVTGAFRDLGIDSLTAVELRNLGLAKVTGLR 1585
Qy 39 ----- 38
Db 1586 LPATLVFDYTPPAVLARLGEFTGENPVLVRTASVVGODEPLAIVGMACRUPGGVSSPE 1645
Qy 39 ----- 38
Db 1646 DLWRLVESGTDAISGFPADRGWDAESLFPDPAVGKVCYVEGGFLDSAAESFDAGFFGIS 1705

Qy 39 ----- 38
Db 1706 PREALAMDPOORLIMEVSWEAFERAGIEPGSVRGSDTGVFMGAYAGGYGAGADLGGFAAT 1765
Qy 39 ----- 38
Db 1766 ASATSVLSGRSVYFFGLGEPAITVDTACSSSLVALYQAGYALROGECSLALVGCVTMAT 1825
Qy 39 ----- 38
Db 1826 POSFVEFSRKSGLDSDGRCKAFADSADGTGWAESGVGLLVERLSDAQAKGHQVLAVVRSS 1885
Qy 39 ----- 38
Db 1886 AVNQDGASNGLSAPNGSPSQOVIRQALANAGLTAEVDDVVEAHGTGTLGDPIDEAQAVIA 1945
Qy 39 ----- 38
Db 1946 TYQDRERPLLGLSKSNIGHAQAAGSVGIKVMALQHNTVPRTLHVDEPSRHDWAA 2005
Qy 39 ----- 38
Db 2006 GAVELVRENQWPCTDRPRRAGVSSFCVSGTNAHVILESAPPAQAEAEQVETPVVASD 2065
Qy 39 ----- 38
Db 2066 VLPLVISAKTOPALTEHEDRLRAYLAASPGVDTRAVASTLAIVTRSVFEHRAVLGGDTVT 2125
Qy 39 ----- 38
Db 2126 GTAVSDPRVVPFGQGMQWLGMSALRDSIVFAERMAECAPALRPFVMDLFTVLDDP 2185
Qy 39 ----- 38
Db 2186 AVDRVDVQPASWRMVSLLAAVMQAAGVRPDAVIGHISQGEIAACVAGAVSMRDAARIV 2245
Qy 39 ----- 38
Db 2246 TLRSEIARGLAGRAMASVALPAQDVVELVDGAMIAAHNGPASTVIAGTPEAVDHLTAH 2305
Qy 39 -----ELLRD----- 43
Db 2306 EARGVRRITVDYASHTPHVELIRDELIDITSSSSQAPVVPWLSTVDGWSVDSPLDVE 2365
Qy 44 ----- 43
Db 2366 YWYRNLRPEVGFHPAVGQLOAEGDTVFVEVSASPVLLQAMDDVVVTVATLRRDDGATRM 2425
Qy 44 -----P-SLG--- 47
Db 2426 LTALAAQAVVHGVTVDPAILGTATTTRVLDLPTAFQHORYWLRSDRAAADGHPLLGTW 2485
Qy 48 ----- 47
Db 2486 DLPASDGVVLTGRVSLATHTWLAHVRGVSLLPGTAPVDLVVRAADEVSDVDDELVIE 2545
Qy 48 ----- 47
Db 2546 TPLLLPQTGGVQLSVSVCGADESGHRAVMVFSQADNTDTWTRHVTATVSTSDSTVSLPEF 2605
Qy 48 -----AQ 49
Db 2606 ASWPPAQARPVSVADFYDRLAAAGTEYGAFOGLQAARWDRDGTVYAEVYLAEBOEAAR 2665
Qy 50 FRVH----- 53
Db 2666 FAVHPALLDAAMHASVLHTPDTDOQSVRMPFSWNHVQIRATDTAMLRAVAATPTDGSVR 2725
Qy 54 ----- 53
Db 2726 VADDTGRPVATIGSLVTRPVATDLTGSAAADDLLTLVWTEIPTPQSSLSVGRVEDLADQ 2785

Db	4946	AICDVSDRAGLARI	LAGVSPHEPLTAVIHTAGALDDGVVESLTAQOLETVLRPKADGAWH	5005
Qy	91	-----	-----	90
Db	5006	LHELTRDADLAA	FVMYSSAAGVLGSGGQGYAAANAFDALAEQRGGGLPALAVANGWL	5065
Qy	91	-----	-----	90
Db	5066	EDASGLTAEMTDT	DRDRIRRGCLRAISAGHGMGLLDAASRHGEPVLLAAAMEPVREAEVP	5125
Qy	91	-----	-----	90
Db	5126	ALLRLLHRPARRAA	STGSSVQWLARLAPVEREKALLKLVCDGAATVLGHADASTIPAT	5185
Qy	91	-----	-----	90
Db	5186	AAFKDLGIDSLTAV	ELRNSLTAKTLRLPATLVFDYPTTALAARLGEWFGETVPVVRT	5245
Qy	91	-----	-----	90
Db	5246	SVSVVAQDEPLAIV	GMACRLPGVSSPEDLWRLLESGLTDAVSGFPTDRGMDVENLFGPAA	5305
Qy	91	-----	-----	90
Db	5306	GDSYRLQGGFLDAA	GFDAFFGIGSPREALAMPQORLLVLEVSWEAFERAGIEPGSVRGT	5365
Qy	91	-----	-----	90
Db	5366	DTGVFMGAYPGYG	IGADLGGFATASAVSVLSGRVSYFFGLEGPATVDTACSSSLVAL	5425
Qy	91	-----	-----	90
Db	5426	HQAGYALROGECS	LALVGGVTYMATPQTVEFARQGLAGDRSKAFADSADGAGFSEGV	5485
Qy	91	-----	-----	90
Db	5486	GULLVERLSDAQK	HQVLMRLSSAVNQDGASNGLTAPNPGSQORVIOAALSNAAGLAH	5545
Qy	91	-----	-----	90
Db	5546	EVDVVEAHGTGTL	GDPIEAQALLATYGODREQPLLLGSKVSNLGHQTQAAAGVGIKMW	5605
Qy	91	-----	-----	90
Db	5606	MALQRFVPRTLHV	DEPSRHDWSAGAVALTENQWPMQMDGRRRAGVSSFGISGNTAHV	5665
Qy	91	-----	-----	90
Db	5666	ILESAPPTQADNAV	IERAPEWLPMVISARTQSALTEHEGRLRAYLAASPGVDMRAVAST	5725
Qy	91	-----	-----	90
Db	5726	LAITRSVFEHRAVL	LGDDTVTGTATDPRVVFVPGQSORAGMBEELAAAFVVFARIHQ	5785
Qy	91	-----	-----	90
Db	5786	QVWDLLDVDPLEVN	ETGYAQPALFALQVALFGLLESKWVRPDAVVCHSVGELAAAGVSL	5845
Qy	91	-----	-----	90
Db	5846	WSLEDACTLVSAR	ARLMQALPAGGVMMVAVPVSEDEARAVLGEVETAAVNGPSSVVLSDG	5905
Qy	91	-----	-----	90
Db	5906	EAAVLQAAEGLKRW	TRLATSHAFHSARMEPMLEEFRAVAEGLTYRTPQVMSAAGDQLTIT	5965
Qy	91	-----	-----	90
Db	5966	EYVWRQVROTVRGE	QVASYEDAVFVELGADRSLARLVDGVAMLHGDHEAQAASALAH	6025
Qy	91	-----	-----	90
Db	6026	YVNGVTVDWPA	LLGDAPATRVLDLPTYAFQHQRYWLEGTDRMAAGGHPLLGEAVAVPGSD	6085
Qy	91	-----	-----	90
Db	6086	GVLITGRVSLATH	PWLADHAVRGSVLLPGTGFVELVVAADVEGCDVIDELVIETPLLP	6145
Qy	91	-----	-----	90
Db	6146	QTGGVQLSVVAES	DDSGRRRAVTVFSRADNVDTWTRHATVTVNSDITLTLSDLSAMPPA	6205
Qy	91	-----	-----	90
Db	6206	QAQPVNVADFYDQ	LTVVAYEYGFQGLQAARWGDVTHAEVALAEBOAQDAARNVHPA	6265
Qy	91	-----	-----	90
Db	6266	LLDAALHAGILNT	PDTEQGVRLPFSMNQOVHVAVSAMLRVAVTQAADGMSVRVADDTGR	6325
Qy	91	-----	-----	90
Db	6326	PVATIGSLVTRPV	TADTLGSTDLLALTWTEIPAPOETSITICORFEDLADADVPEVV	6385
Qy	91	-----	-----	90
Db	6386	VFTARPDSSENPL	AQTRTLTAQVLQKIQAWLAEBERFTDSTLIVRIGTCLAAAVSGLMRS	6445
Qy	91	-----	-----	93
Db	6446	QOSEHPGRFVLVE	SDDDTLAPDQLAATVELDEPRLRISDGRFEVPRLTRTHAABPESEKV	6505
Qy	94	-----	-----	93
Db	6506	WDPDGTVLITGSG	VLAGIAARHLVAERGVNRHLLLSRSPDGALIISELSELGAQOVATAV	6565
Qy	94	-----	-----	93
Db	6566	CDVSDRPGELVR	LADVSPHEPLTAVIHTAGVVDGVVESLTAQRLDVTLRPKADGAWHLH	6625
Qy	94	-----	-----	93
Db	6626	ELTRDADLAAFM	YSSAAGVFGSAGOGNYAVANFLDALAEORRAEGLPALALAWLWEG	6685
Qy	94	-----	-----	95
Db	6686	TSGLTANLTDH	ORIRSRGMRAISAEHGMRLFDGASRRRDPVLVAAMEPVREAEVPAM	6745
Qy	96	-----	-----	100
Db	6746	LRSLHRPVARRAA	SGGAVRCLAAALAPBEERAKALVKVWCDSAAATVLGHADVDSIPVTAAPR	6805
Qy	101	-----	-----	100
Db	6806	DLGVDSLTAVEL	RNSLTAKTGLRLPATLVFDYPTPGALAAARLEELFTGENPVQVTRPVSA	6865
Qy	101	-----	-----	100
Db	6866	VQDEPLAIVGMAC	RLPGGVSSPEDLWRLVESGTDASISGFTDRGMDVENLFDSDPDAA	6925
Qy	101	-----	-----	100
Db	6926	KSYCVEGGFLATA	NFDFASFFGISPREALAMDPOORLVLEVSWEAERAGIEPGSVRGSD	6985
Qy	101	-----	-----	100
Db	6986	TGVFMGAFPGG	YGIGADLEGYGATAGNLVLSRLSYFFLEGPAVTVDTACSSSILVALHO	7045
Qy	101	-----	-----	101
Db	7046	AGYALROGECS	LALIGGVTVMATPHTFVFSRQRGLASDGRCKAFADSADGTGWSEGVGV	7105
Qy	102	LVITRF	-----	107
Db	7106	LLVERLSDAQKH	QVQLAVVRSSAVNQDASNGLSAPNGPQORVIRQALANAGLTAEV	7165

Qy	108	-----	107
Db	7166	DVVEAHGTGTTGLDPIEAQVIATYQDRDQPVLLGSKSVNGHTQAAAAGVSGVIKMYMA	7225
Qy	108	-----	107
Db	7226	LOHGLVPTLHVDEPSRHWDTGAVELVTENQSWPEACRPRAGVSFVSGTNAHVIL	7285
Qy	108	-----	107
Db	7286	ESAPPTQAVDDVRPADAPVVASVMASELVLVISAKTQSALAEYEGRLRAYLAASPGVDM	7345
Qy	108	-----	107
Db	7346	RAVASTLAMTRSVFEHRAVIGDVTYSGAATDPRVVFVPOGQSORACMGAEALAAFPV	7405
Qy	108	-----	110
Db	7406	FARIHOQVWDLDDVPLEVNETGYAOPALFALQVALFGLLESWGVPRDAVIGHVSGELAA	7465
Qy	111	-----	115
Db	7466	AYVGLWSLEDACLTYSARARLWQALPAGGVMAVAVPVSEDEARAVLGEVGEIAAVNGPSS	7525
Qy	116	-----	115
Db	7526	VVLSGDEAAVLQAAEGLKWTRLATSHAFHSARMEPMBEEFRAVAOGLTYHAPGVVNAAG	7585
Qy	116	-----	119
Db	7586	DRVMTAEYWRQVRDVTFRGEQVASYEDAVFVELGADRSRLARLVGVMHLGHDHETQAAI	7645
Qy	120	-----	122
Db	7646	GALAHVLVNGVTVDWTALLGDVPVTRVLDLPTVAFQOQRYWAEVGRSADVSGAGLDAVGH	7705
Qy	123	-----	122
Db	7706	PLIGAVVALPGSGVLLTGRVSLATHAWLADHAVRGSVLLPGTGFVELVVRAADEVGCDV	7765
Qy	123	-----	122
Db	7766	VDELIVEAPLLLPTQGGVQSVSVGEADESGHRVVTIFSRADNADTWVRHVSAVISTSDS	7825
Qy	123	-----	122
Db	7826	TVSLSEFASMPQAQPVDTGFDPLMAEGYEYGPAPQGLQAARWYGDTVFAEVALAE	7885
Qy	123	-----	122
Db	7886	QVREAAARYPSVHPALLDAAHACTLECADAEVGVGLPFPWDGVRVHAGGSANLVRVATQA	7945
Qy	123	-----	122
Db	7946	ADGWSVRVADEIGRPVANVGSVLTVPYADALGSAADDLLALTWAEIPAEGTGLTVGRF	8005
Qy	123	-----	122
Db	8006	EELVSDGDPVPEVAVFTALPDNDPDDPLEQTRKLTQVQLQAVQEWLGGERSFSDTLVVRT	8065
Qy	123	-----	122
Db	8066	GTGLAAAGVSGLMRSAQSEHPGRFVLVESDDDLALDQALAAVGLDEPRLISDGRFEAP	8125
Qy	123	-----	122
Db	8126	RLTRTHAAPESEKYWDPDGTVLTITGGSVLAGIARVHLVTERGVRVHLLLSRSAPDEAL	8185
Qy	123	-----	125
Db	8186	INQLGELGARVETAACDVSDRAALAQVLAGVSPHEPLTAVIHTAGVLDDGVVLESITAORL	8245
Qy	126	-----	125

Db	8246	DAVLRPKADGAWNHLHELTRDADLAAAFVMYSSAAGVLGSGCGNYAAANAFDALABORHA	8305
Qy	126	-----	125
Db	8306	EGLPALAVAWGLMEDASGLTAQLTDTDRDIRRGGLRAISAHEGMGLFDSASRHSSEPLV	8365
Qy	126	-----	125
Db	8366	AAPMEPVDAEVPALLRSLHRPVARRAAAAGGARWLAALAPAREKALLKLVSDGAATVL	8425
Qy	126	-----	125
Db	8426	GHADTSTIPATTAFKDLGIDSLTAVELRNSLAKATELRLPATLVDFYPTPTTALAARLDEL	8485
Qy	126	-----	125
Db	8486	FTGENPAPVRGPVSAVAQDEPLAIVGMACRLPGGVSSPEDLWRLVSGTDAISGFTDRG	8545
Qy	126	-----	125
Db	8546	WDVENLYDPDPAKPSKSVYVGGFLDAAAGFDASFFGISPREALAMPQORLMLEVSWEA	8605
Qy	126	-----	125
Db	8606	FERAGIEPGSVRGSDTVGICAYPGYGIGADLGGFGTTAGAAASVLSGRVSYFFGLEGA	8665
Qy	126	-----	131
Db	8666	FTVDTACSSSLVALHQAGYALROGECSLALVGGVTVMPTPQTTFVFFSRQGLSADGRCKA	8725
Qy	132	-----	131
Db	8726	FADAADGTWAEVGVLLVERLSDAQANGHQILAVVRSSAVNQDGASGLSAPGPSQOR	8785
Qy	132	-----	131
Db	8786	VIRAAALSNAGLAPHEVDVVEAHGTGTTGLDPIEAQVIATYQGGREPELGLSLKSNVGH	8845
Qy	132	-----	133
Db	8846	TQAAAGVSGVIXKMWALQHSMPRTLHVDEPSRHWVDSAGAVELVAENQPMETGRPRA	8905
Qy	134	-----	133
Db	8906	GVSSFGISGTNAHVILESAPOSGDGTAGTPELVSELVPLVISAKTOPALTEHEDRLRA	8965
Qy	134	-----	133
Db	8966	YLAASPGVDIRAVASTLAVTRSVFEHRAVLLGDETVTGTAVSDPRIVFVFGQGHQWLGM	9025
Qy	134	-----	133
Db	9026	GSAIRDSSVVFAERMAECAALSEFVDWDLFAVLDDPAVVDVVDVVPASWAMVMSLAAV	9085
Qy	134	-----	135
Db	9086	WQAAGVPRDAVIGHSGEIAAACVAGAVSMRDAARIVTLRSQAIARGLAGRGAMASVALP	9145
Qy	136	-----	135
Db	9146	AQDVLDGAWIAAHNGPASTVIAGTPEAVDHLVTALRORGAGAADHVDYASHTPHVELI	9205
Qy	136	-----	139
Db	9206	RDELLOITSSSSQDPLVPLWSTVDTGTVWDSPLDGEWYRNLRKREPVGPHPAVSQLOAQGD	9265
Qy	140	-----	139
Db	9266	TVFEVSASPVLMQAMDDDDVTVTLRRDDGDATRLMTALAAQYVHVHTVDWRAVLGDVP	9325
Qy	140	-----	139

```
Db 9326 ATRVLDLPTAFOHQRYWAEGRSADVSAGLDVAUGHPLLGAVLAMPGGDGVMLTGRVSL 9385
Qy 140 -----TGF----- 142
Db 9386 ATHAWLADHAVRGSVLLPGTGFVVLVRAADEVACDVVDDELIVEAPLLLPQTGGVQLSVS 9445
Qy 143 ----- 142
Db 9446 VGEADESGHRAVTVFSRADSDAWVRHVSATVSVSDTTVPTSDLTAMPPAQAKPVDVAGF 9505
Qy 143 ----- 142
Db 9506 YDQLTRAGYEGPAFGCLOQAARWGDGTVAEVALAEEOQDAARFAVHPAVLDAAHAGI 9565
Qy 143 ----- 142
Db 9566 LNTPDADRTVRLPFSNNHVQHVHTGSATLRVATRVADGWGRVADDIGRPVATIGSLV 9625
Qy 143 ----- 142
Db 9626 TRPVAADALGSAVDLFTLTWTEIPVSQQGVTVGKFEDLADGEVMPDVMVVFTALPDSG 9685
Qy 143 ----- 142
Db 9686 DPLAQTRRLTAEVLOAVQVWLAGERTDSTLVVRTGTGLAAAAVSGLMRSAQSEHPGRV 9745
Qy 143 ----- 142
Db 9746 LVESDDTLTPQLAATVGLDEPLRLVIDGRVEAPRLTRTGVAEPEGEVDPDGTVLIT 9805
Qy 143 ----- 142
Db 9806 GGSGLAGIAARHLVAERGVRHLLLSRSPADEALISELAELGAAVVDVAVCDVSDRAGL 9865
Qy 143 ----- 142
Db 9866 ARVLGVSPDPLTAVIHTAGVLDGQVVESTARRLDTVLRPKAGCANHLHETRDIDLA 9925
Qy 143 ----- 142
Db 9926 AFVMYSSAAGVLGSAGOGNYAVANAFVDALAEQRAEGLPALALAWGLWEDASGLTAKLT 9985
Qy 143 ----- 142
Db 9986 GTDHRIRRGSLRTITAERGMLFDIASRQGEPLVATPMPREVVEVPALLRLLRPVA 10045
Qy 143 -----DLGV 146
Db 10046 RRAASTGSSAQLVGLAPEERAKALLKVVRDASAATVLGHADARSIPATGAFKDLGV 10102

RESULT 12
O85168
ID O85168 PRELIMINARY; PRT: 9376 AA.
AC O85168;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Stryptomycin synthetase.
GN SYR.
OS Pseudomonas syringae (pv. syringae).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=321;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99047670; PubMed=9830033;
RA Guenzi E., Galli G., Ggurina I., Gross D.C., Grandi G.;
RT "Characterization of the syringomycin synthetase gene cluster. A link
RT between prokaryotic and eukaryotic peptide synthetases.";
RL J. Biol. Chem. 273:32857-32863(1998).
DR EMBL; AF047828; AAC80285.1; -.
DR HSSP; P14687; 1AMU.
```

```
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Ppantne_attach.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF00501; AMP-binding; 8.
DR Pfam; PF00668; Condensation; 9.
DR Pfam; PF00550; pp-binding; 9.
DR Pfam; PF00975; Thioesterase; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS50075; ACP DOMAIN; 9.
DR PROSITE; PS00455; ACP BINDING; 8.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_5.
KW phosphopantetheine.
SQ SEQUENCE 9376 AA; 1029826 MW; F770C08975EF9CES CRC64;

Query Match 30.7%; Score 237; DB 2; Length 9376;
Best Local Similarity 1.2%; Pred. No. 27;
Matches 94; Conservative 24; Mismatches 28; Indels 7809; Gaps 29;

Qy 1 AAGG----- 4
Db 980 AAGGSEVVVPANLIPEHCDRITPQMLPLINLSOAQIDHVVKMPGGVANVDIYPLAPL 1039
Qy 5 ----- 4
Db 1040 QAGILYHHISAEQGPYTLKALFALSADRAQLDDPSGALQGVINRHDILRTAVLWEDLDEP 1099
Qy 5 ----- 4
Db 1100 VQVVLRAQLHVTELLDPDGPVDEQLHQRDRRHRYLDVRTAPLMRIVFSDHPVNDRW 1159
Qy 5 ----- 4
Db 1160 LAMLLCHLVSDATSLVLOEIOAHLLGQGNALGEAVPYRNYVAQRLGVSEAQHEAFP 1219
Qy 5 -----ILHL-- 8
Db 1220 RDMLGDIDEPTLPFGLQDVQDRGRNLEEASVTLAAELNLRARQAGVSVASLHMLAW 1279
Qy 9 ----- 8
Db 1280 ARVLGVNSACEQVVGTVLLGRMQADGADRALGMFINTPLRVDIGATTVAEGLKATHE 1339
Qy 9 ----- 8
Db 1340 RLTAALLGHEAPLVLAQRCSGVAAPTPLFSALLNRYHSVSDVRQKLPGLGSSLRHRRP 1399
Qy 9 ----- 8
Db 1400 YQLPVDNLNVDLGDRAFTLTITQAVQOISATRIGEYMQVALRNLYDALEHTPQALNSLSIL 1459
Qy 9 ----- 8
Db 1460 PDDERELLLTGFNDTAHPYPRDVLHQLIEQAAAQRPDACAVRSGSGTLLTYAELNQAN 1519
Qy 9 -----ELLVA-- 13
Db 1520 QLAHRLIELGVEPDRVAVSRRGAEVMVALLGILKAGGAYVDPDLPDARQAYMLSDS 1579
Qy 14 ----- 13
Db 1580 APRAVLTSHELLADLPDLGVPAVLVDGRDSDSALLKKQPTGNPDAKALDLPNHLAYVLYT 1639
Qy 14 ----- 13
Db 1640 SGSTGTPKGVNNEHLGVNRLHARDAYQVNSQDRVLQKTPFGFDVSVWEFFLPILTCAE 1699
Qy 14 ----- 13
Db 1700 LVMARPSGHQDPDYLAQVSDAGITLLHFVPSMLDVFLEHRSTRDFFPOLRRVLCSEALP 1759
Qy 14 ----- 13
```

Db 1760 RALORRFEQHLKGVELHNLGYPTAAIDVTAWECRPTDGDSPICRPIANIQMHVLDAL 1819
Qy 14 ----- 13
Db 1820 GOLQPMGVAGELHIGGIVARGYLNOPQLSAERFIADPFSNDPQARLYKTGDVGRWLANG 1879
Qy 14 ----- 13
Db 1880 ALEYLGRNDFQVKIRGLRIEIGEIEAALAKHPAVHEAVTAREDPGDKRLVAYYTQAE 1939
Qy 14 ----- 13
Db 1940 HTAVDLETLRGLHQVLPEYMPAIYVLEAMPLTSNGKLDKALPAPGDALISRGVEA 1999
Qy 14 ----- 13
Db 2000 POGEEIQAVIWDLLGVEQVGRHDFHFGHSHLLAVSLIDRLRKQDLNLTNTVFTA 2059
Qy 14 ----- 13
Db 2060 PSVREMAIAISOEKOALFRAPANRIPAHCTQLTPDMPLVLVELSVAQIELIASAVPGGAAN 2119
Qy 14 ----- 13
Db 2120 IQHIYPLAPLOEGILFHYLLNERDAYLVRSTIEFDSRARLDAFLEGLQTVIDRHVLR 2179
Qy 14 ----- 13
Db 2180 SVHWGLPQAVOVVHROAQSPHITLRLPDEDALSOLDRLSDPGRRLDLRQAPLLLAYI 2239
Qy 14 ----- 13
Db 2240 ARDPDSERWLLALIDHHMISDHVTLEILEIRLLMRGOSAEILLPPQYREFVAQTASP 2299
Qy 14 ----- 13
Db 2300 SSAHEAFTGRADVDSPTAPFELLEVOGDNDVEESELSSDLCAIRTOARQGMSP 2359
Qy 14 ----- 13
Db 2360 AVLPHVAWAQVLARCTGRDDVVFGTAVTGRLOGLGAERANGMFMTLPVRVOLATOSVQ 2419
Qy 14 ----- 13
Db 2420 ELVMATHRDLSELLSHEQASIALAQRCSVATGVPLFSSLLNYRHQEDSOLQWPGRL 2479
Qy 14 ----- 13
Db 2480 DSAERTNYPLCLSVNDYGSIDLGLIHSVQPADPORLCAAMQCALEQLTDALHTPQKEVT 2539
Qy 14 ----- 13
Db 2540 QLDVLPAAERNLLLETFTNQTRQDYPDTLCIHLFEAQVTRQPDIAIAVFOAQRLSYAEIN 2599
Qy 14 -----VGPD----- 17
Db 2600 ROANRLAHLILGLIGDPPDRAICVERGEMVGLLGLVKGAGAYVPLDPAYPAERLAYM 2659
Qy 18 ----- 17
Db 2660 IEDSQAALMTQRHLOEYPLTLPLVLDDDKRTFTERDDNPVVEALGVRNLAYVIYT 2719
Qy 18 ----- 17
Db 2720 SGSTGNPKGVMIEHRGLVNYSDAARLFDLSPTDVLQONTLNFDLSVEEIPFALLAGAT 2779
Qy 18 ----- 17
Db 2780 LTPSREIFGSEGTEHNGINPTVLHLTAHWHTLVAEWHKQPVAEQRLQHVRLINVTGDA 2839
Qy 18 ----- 17
Db 2840 LSAOKLWDEVRPAHTRLLNTYGPTEATVSTAAVYSHDAAAGSESGNATIGKPMANT 2899

Qy 18 ---VFOAHQE-----D 25
Db 2900 RIVLLDAHQQPVPYGVAGEIYIGDGVARGYLNLEEVNAERFLADPFSSEPDARMYKTGD 2959
Qy 26 TERYV----- 30
Db 2960 LARYMADGRIEYLGRRNDFQVKRGFRIELGEIARLGNCTGVKEAVVIAREDNPGDKRLV 3019
Qy 31 -----LT----- 32
Db 3020 AYVVAOQSQLTAADLRAELAPQLAEMYMLPSAFVLLDELPLTPNRKLDKRALPAPDADAL 3079
Qy 33 -----NLN 35
Db 3080 ISRGYEAPOGETETQIVAIWQDLLGIEQVGRHDFHFGHSHLLAVSLIERLURKQGLNIN 3139
Qy 36 I----- 36
Db 3140 VKTVFTAPTREMALAIISODKHVLFQVPANRIPAHCTQLTPDMLPLVELSVAQIELITSA 3199
Qy 37 ----- 36
Db 3200 VAGGAANIQDIYPLAPLODGILFHYLLNRERDAYLMRSMIEFDSRARLDAFLEGLQTVID 3259
Qy 37 ----- 36
Db 3260 RHDILRSSVHMIGLPOAVOVVHROAQLPVHTLTLPTEEDALSOLDRLSDPGRRLDLRQA 3319
Qy 37 ----- 36
Db 3320 PLLLAYIARDPNSERWLLALIDHHMISDHVTVELILEIRLLMRGOSADLLPPQYRDFV 3379
Qy 37 ----- 36
Db 3380 AQTLASPSSAHEAYFTRRLADVDSPTAPFELLEVOGDNDVEEAKLNSDLCAIRITQA 3439
Qy 37 -----GAE----- 39
Db 3440 RERGHSPAVLFHVAWAQVMARCTGRDDVVFGTVTGRLOGTAGABRAMGFMFMTLPVRVQ 3499
Qy 40 ----- 39
Db 3500 LTTQGAQELVMATHRDLSELLSHEQASIALAQRCSVATGVPLFSSLLNYRHQEDNRLO 3559
Qy 40 ----- 39
Db 3560 WPGMRLDGTERTNYPLCLSVNDYGSSELDLIHSHQWAPANPORLCAAMQCALEQLTDALAH 3619
Qy 40 ----- 39
Db 3620 TQMAVTOLDVLPAAERNLLLETFTNQTRQDYPDTLCIHLFEAQVTRQPDIAIAVAVOQOR 3679
Qy 40 ----- 39
Db 3680 LSYADLNQANRLAHLILSLGIVPDDRVAICVERGEMVGLLGLVKGAGAYVPLDPAYP 3739
Qy 40 -----LRLD-----PSLGAQ- 49
Db 3740 AERLAYMITDSQAALLTLPLGLQDRLPALSMPLVLLDDEQVGLAECDNDNPVPTLGVNR 3799
Qy 50 ----- 49
Db 3800 LAYVIYTSGSTGNPKGVMIEHRGLVNYSDAARLFDLSQSDTVLQONTLNFDLSVEEIPF 3859
Qy 50 ----- 49
Db 3860 ALLAGATLAPSREIFGSEGTEHIGIQTPLVHLTTAHWHTLVAEWHNQPVAAEQRLQHVRL 3919
Qy 50 ----- 49
Db 3920 INVTGDALSOKLWDEVRPAHTRLLNTYGPTEATVSTAAVYSDAAAGSESGNATI 3979

QY 50 ----- 49
Db 3980 GKPMANTRIYLLDAHQQPPYGVAGEIYIGGDCVARGYNLEEVNAERFLADPFSESDA 4039
QY 50 ----- VKMVL----- 60
Db 4040 RMYKTGLARYWADGRIEYLGRRNDFQVKVGFRIELGETEARLGNCKGVKEAVVIAREDN 4099
QY 61 ----- 60
Db 4100 PGEKRFVAVVQPOQTITAAELRAELAPOLAEYMLPSAFVLLDELPLTPNKKFDRKALP 4159
QY 61 ----- TEPEGANIT-ANLTSSLSV----- 80
Db 4160 APADDAFASREHVEFGATEIALAQIWSLLDLERVRHDOFFELGCHSLLAMRLISQAR 4219
QY 81 ----- 80
Db 4220 HQGVELGLAALFAHPEVSALAVATAQAGRSKFPDIPVARDQAWPLSFGQORLWFLAQM 4279
QY 81 ----- 80
Db 4280 EGASAAHYHIPAGLSLHGNLNLKALQALERIVARHEGLRTTFWQGGDQGPVORISPADTG 4339
QY 81 ----- 80
Db 4340 FNLQMHDLQGLADAEKLOALASEESLSQFDLQOQPLIRGRRI RMAEDHHVLLTLHHIV 4399
QY 81 -CWS----- 84
Db 4400 SDGWSVDVLTRELSALYAFSODDDPLAPLELOYLDYAVWQRRMLSGDVLOQQSNFWQO 4459
QY 85 ----- 84
Db 4460 TLADAPALLMLPTDRARPALQYAGAAALPVVFXDLTRGLKALSRRGSTLFWTVMAAWA 4519
QY 85 ----- 84
Db 4520 GLLGRLAGQDDVVIGTPVANRRTRSEVEGLVGLFVNTLAI RVDLSDKPTAETLLARVKQT 4579
QY 85 ----- OTINP----- 89
Db 4580 LDAQAHQDLPFQEVVEVINPLRSLSHSPIFOAMLSWENNEASDLTLGDMTLKSLIADT 4639
QY 90 ----- 89
Db 4640 AQFDLTDMAEVDQLVGTLEYATALFDESTMRRYLYGFQVLEAMVADDDQVLEHVALL 4699
QY 90 ----- 89
Db 4700 GADEREHLLAGLNATEAPFPQDRTIHQLFEERVQAQDAIATAVAFGAQRLSYAELNRQANR 4759
QY 90 ----- EDDT 93
Db 4760 VAHHLISLGIKPDRVAICVERGVEMLIGVLGVKAGAAVPLDPAYPAERLAYMIEDST 4819
QY 94 ----- 93
Db 4820 PSALLAQRDVQAHPLTLDLPLVLLDEQRTTLSEDDNPVVTLAGHNLAYIYITSGTG 4879
QY 94 ----- 93
Db 4880 VPKGVVHEHRGLFAVSAANEQLYALHAPLNHLQWAGFFDVFSADLIRSLAFGGTLCVLCPR 4939
QY 94 ----- DP----- GHADLV----- 102
Db 4940 ETLMDDPALYRLLEEISIGFADFVLAVNALLGWEEETHGDLFSMRTVVCSDIWTASHA 4999
QY 103 ----- 102
Db 5000 ROLRKLCDGHVVQVQAYGVTEASIDSTCFEATSQVDGVLPIGRALANTRIYLLDELQO 5059
QY 103 ----- 102

Db 5060 AVPGVAGELYIGGAGIARGYLNLPOLTAERFMDNPFVAGERLYRSCDMARYRADGNIEF 5119
QY 103 ----- YITR----- 106
Db 5120 LGRNDSQAKLURLLELGEIEARLAEVAGVRESLVVIREDSGGTPPKLIAYFEVATRDES 5179
QY 107 ----- 106
Db 5180 GPALTPRALRQQLNLPEYMIPAAFVRMAALPLSANGKLDLRRALPEPDADAFDQHDFA 5239
QY 107 ----- 106
Db 5240 ADGPLETAIAIADVLGVAQVGRHDDFFALGGHSLLMVRVLAOVROOLNLEVSFSVFFA 5299
QY 107 ----- 106
Db 5300 APVURQFAERLGNTOQNARVAIKPVQRSGALPLSYAQORLWFLAQLEGSSAAHYHIPAGLR 5359
QY 107 ----- 106
Db 5360 LRGNLQASLQALDRIVARHEALRTTFVQEQGPAPQRISAETGFRLOQLQVLAGOTDA 5419
QY 107 ----- 106
Db 5420 EDKLLATAAQEASEHFDLVNGPLVRGRLVNRANDNHVLLVTMHHILSDGWSGTGVNLREL 5479
QY 107 ----- 106
Db 5480 ALYAPFROGAGDPLPALPVQYVYDALWQDMLSGDVLOOQRYWQOALAGAPALLTLPTD 5539
QY 107 ----- 106
Db 5540 RPRPAQDYSQOTLELVDTLTQTRGLKALSQRHGSSLFMTVMGAWAALLGRLSGQDDVVI 5599
QY 107 ----- 106
Db 5600 GTPVANRMSAEVENLIGFFVNTLAI RVDLSGTPSVEALVRQVKORTLAAQANQDLPFEQV 5659
QY 107 ----- FDLE----- 110
Db 5660 VEVVOPQRSLSHSPITFOAMLSQNNEDTALVGLDITLQGVAVAGDTAKFDLALDIDGEVDG 5719
QY 111 ----- 110
Db 5720 QLIGTLEYATALFDESTMRRYRGYFLRLLEAMADDDQVLEQVPLLDTAEREYLLKDINA 5779
QY 111 ----- 110
Db 5780 TERTYPVQQLMHLRLEFAHAEAAPOAIAVRQGEQTLTYAELDSRANALAHQHLRKHGVGPGT 5839
QY 111 ----- 110
Db 5840 RVAIILDRSVELLASMLATLKCGAAYLALDRLAPEERLRFMLEDESEAIMLLSRSDLTAPD 5899
QY 111 ----- 110
Db 5900 MTPRLDLDLLELSALNQEPVVPATEVAGETPACIIYTSGSTGVPKGVIVTHNGI VRLVQD 5959
QY 111 ----- 110
Db 5960 NGYDYSABDRVAFSSNPAPFADASTPEIWGALLNGQSVIIEPQVILLEVPAFALLKRHGV 6019
QY 111 ----- 110
Db 6020 TAMISSTALFNLYAGLIPALAGLRWIMCGGERADPASFRRVREHSAQVRLFNGYGPTEG 6079
QY 111 ----- 110
Db 6080 TTCATSYEIFDVLPTDLSLPIGKPNANVRVYVLDARREPVPNGVGDIIYIGGTGVALGYL 6139
QY 111 ----- LPDGNRQ-----VRG----- 120
||| :
|||

Db 6140 NRPELTAERSEDFPSQOAGARLYRTGDLARWLPGDNLEYLARNDGOVKYGRFVELGEI 6199
QY 121 ----- 120
Db 6200 ESVLHLYDGYRNSVVAHAESPGDTRLVAYTYVHAGVAAPDPEDLRAQLSADLAEMYWPS 6259
QY 121 ----- 120
Db 6260 VFVRLDALPLTNKGVQDKALPVPDLNAMFERTYVAPEGATEQALAEI FOELLGLSERVGR 6319
QY 121 ----- 120
Db 6320 HDGFFELGCHSLLAAQLVSRVROQLNGDMALROLFNHPTVAELAKVVDGLQTVDSDSIEP 6379
QY 121 ----- 120
Db 6380 IERNAPLALSFSQORLWFLDRLDPGASSAYHMPMSLLRGELDHRAKKAALDRLVARHES 6439
QY 121 ----- 120
Db 6440 LRTTFELHGEQPVQVIAADSGFALAEDDLRSQPYEQASLNASRIADSEAAAPFDLRQGP 6499
QY 121 ----- VTO----- 123
Db 6500 LIRGRLLRLADDEHMLLIOTHHIISDGWSVGVLINEFTALYQAFTEQRPDPLALSIOYA 6559
QY 124 ----- LGACST----- 131
Db 6560 DYAAQORRTFTGERLAEQADLWREHLGA - PTLLSLPTDRPRPVVQSYRGGAVPVTIDA 6617
QY 132 ----- WSLIT----- 137
Db 6618 ALHQLERFCOAHNVTLFMGLLSAMSVMTRLGNERNWDVWIGVPSANRGTETENLIGFFV 6677
QY 138 ----- 137
Db 6678 NALALRVDLTONPSVAQLLEQVRQTTLAAHEHQDIPFEQVIEALQPPRSMHSPLCOVAL 6737
QY 138 ----- 137
Db 6738 SLDNSTIGGELKPLGLSLHPVLOAHETAQFDLMTLASENGALNGVIEYASDLFDRSTVE 6797
QY 138 ----- 137
Db 6798 RPAQHEHTLEAMVEDVAQVGLPLLSPAQRLASPALLOPKAVFASGLMVHQRFEQFAA 6857
QY 138 ----- 137
Db 6858 AHQONIALVFRHEVSYQALNRRANRLAHELLAOGVRPDDRVAILAERGTMICAVLAVL 6917
QY 138 ----- 137
Db 6918 KSGAAYVPLDPTYPTERLGYLLTDSAPVALLAQSACLDVLPAHSVPVLTLAGDNDSADLA 6977
QY 138 ----- 137
Db 6978 DESVDRNPEPAALGLSREHLAYVIYTSGSTGLPKGVLEHGNVARLPDATAGQNFNGHED 7037
QY 138 ----- 137
Db 7038 WTPFFHSFAPDFSWEIWGALCYGKLVIVPSEVARSPDDFYALVCEQVTVLNQTPSAF 7097
QY 138 ----- 137
Db 7098 RQFIQARERSPOEHALREVWFGGALDFRSLQPWTARTPLSRTRLVNMGYITEITVHATY 7157
QY 138 ----- 137
Db 7158 YPISQSEIDTAMPSLIGPALDDLCLRIIDDYQPPVPVGVNGEYIYGAGVARHYLNRTTEL 7217
QY 138 ----- 137
Db 7218 NAERFIADPYALQSGARLYRTGDVAHYRSOGGVNVNVRNDSQIKIRGFRIELGEIEAQLL 7277

QY 138 ----- 137
Db 7278 ACPEVREAMVILREDRPGDKRLVAYLIAEDGAAPESALLRSQLASVLAEHMLPSAPVTUE 7337
QY 138 ----- 137
Db 7338 TWPLTTNGKLDRAALPAPDQFATVSRDYEAPLGAIEITTLAAAWQELLGVERVGRDHPFE 7397
QY 138 ----- 137
Db 7398 LGHGSFLVISLIERLRQIGLLLDVSTVSAPTLOAAVAVALAGGTAAERVAPANLIPVDC 7457
QY 138 ----- 137
Db 7458 ALTPDMLPLVKLTROELEHIVADVPGGVANVQDIYPLSSLOEGILPHHLLQSEGDAYLMR 7517
QY 138 ----- 137
Db 7518 TLATPDSRALLDKFLCALQVWINRHDIMRSSLRWQGLPOPQVWHRQAOLPVIQLOTAPG 7577
QY 138 ----- 137
Db 7578 EDALQMLRERTNTYHMRDLQOAPLIAAYITYDTRQEKWLMALLDHLHLSDNVTLRLIMG 7637
QY 138 ----- 137
Db 7638 EIQAVMDGRADALPPSPYRNFIAACVQSAEHEAYFRQLLGDVDTTTPAYGVLDVRGG 7697
QY 138 ----- 137
Db 7698 DATILRSVQDLSDDLSARIHSTARAQGVPTSVLFHAANGLVVAATSGRDDGIFGTVLSGR 7757
QY 138 ----- 137
Db 7758 SQOTSGANHALGMFINTLPMRIRLOQNSVRDIVQAYQOLSGLLTHERAPLALACRS 7817
QY 138 ----- EDTG----- 141
Db 7818 DASLPMFTVLNCRHGDVNTAGENIEDMGEOGVHFLGSETRTNYPPIEIAVANRANGF 7877
QY 142 ----- 141
Db 7878 SLTAQSIDGIDPHRTAAAYLQGAELVNALAEODPARLASSLEVIPOAERQLLLDNFDNTA 7937
QY 142 ----- 141
Db 7938 SDFAPAVPIHALPEQVRRNPDAVALVYEDRQLSYRQLNRRANHVARQLQLGVQPDERV 7997
QY 142 ----- 141
Db 7998 AICAERSLDMIAGLLGLVLSKGAAYVPIDPAHPADRMFQMSQPRALLTQGALSPLVGD 8057
QY 142 ----- 141
Db 8058 TPLMLLDSAESLLAADDOAFDANPVVDGLTAENLAVIYTSGSTGOSKGMVHRSVFNF 8117
QY 142 ----- 141
Db 8118 WNVLTRTHQHCHPTPATVALNAGFFEDMSIKGISQLFSGHKLVIIPOLLRANGSELDDL 8177
QY 142 ----- 141
Db 8178 EAHQVHAFDSTPSQDITLLSAGLLERSQYPSVLLGGEAINASTWEKLRNCPTIRLYNM 8237
QY 142 ----- 141
Db 8238 YGPTCTVDATIDLIRDLGKPSIGRPIANQVHVLDARGEPAPLGVAGEIHIGGSVAR 8297
QY 142 ----- 141
Db 8298 GYLNRDELSAERFIVDPFSDAANARLYKTGDLGRWLADGTLEYMGRNDFQVKRGFRIEL 8357

QY 142 ----- 141
 Db 8358 GEIENVLLAVPGIREVVIARNDSDQSQRLLVAVCGESVAAEHLRSELLRHLPYKV 8417
 QY 142 ----- 141
 Db 8418 PSFVOLDALPLTANGKLDRLRALPAGQDALASKVVEAPQGETEBAIEIKWALLHLDQV 8477
 QY 142 ----- 141
 Db 8478 GRNDGFEELGHHLLAVOLLRLRRKLGTRITLRELFDAPTVRGLASLVNAAAPGEAQSI 8537
 QY 142 ----- 141
 Db 8538 PRANRSGRLPLSPSQORLWFLDHLDDHAAAGAAVHLPNALRLTGALDTAALEATLDRLVARH 8597
 QY 142 ----- 141
 Db 8598 ETLTRFELVGEVPQKIAPDSRLPLKQDLRLDLSGERTSTLARLGOENATQLFDLTK 8657
 QY 142 ----- 141
 Db 8658 GPMRLGHLRLVADAHEVLLITLHHIVSDGWSNSVLAEQVSALYAFSQGKQDPLPALPQ 8717
 QY 142 ----- 141
 Db 8718 YVDYAAWQSQDGPALQAIQDFWRKHLECAPSVLNLPLDRPRPAIQSYTGGWVHFSP 8777
 QY 142 ----- 141
 Db 8778 ALSADLRAPSAQGSTPFVLLAGWSMLSHLSEQTQDVVVGTPVANRQHPLEPMIGFEA 8837
 QY 142 ----- 141
 Db 8838 NTLALRVATDRETRNLMDLRKSLTLAAYNHODLPFEQVVSALQPTRNVSHSPLEFQVML 8897
 QY 142 -----FDLGVTI 148
 Db 8898 SLDNTPPSLQLPDLVELLDSEHHTTQFDLSLSL 8932

RESULT 13
 Q9L4W3 PRELIMINARY; PRT; 11096 AA.
 ID Q9L4W3;
 AC Q9L4W3;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE NYS.
 GN NYS.
 OS Streptomyces noursei.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1971;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 11455;
 RX MEDLINE=20334850; PubMed=10873841;
 RA Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
 RA Valla S., Zotchev S.B.;
 RT "Biosynthesis of the polyene antifungal antibiotic nystatin in
 RT Streptomyces noursei ATCC 11455: analysis of the gene cluster and
 RT deduction of the biosynthetic pathway.";
 RL Chem. Biol. 7:395-403(2000).
 RL EMBL; AF263912; AAF71776.1; -;
 DR HSSP; P25715; IMLA.
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR002085; Adh_zn_family.
 DR InterPro; IPR004410; Fabb.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR003880; Pplantn_attach.
 DR Pfam; PF00698; Acyl_transf; 6.
 DR Pfam; PF00107; adh_zinc; 1.

DR Pfam; PF00109; ketoacyl-synt; 6.
 DR Pfam; PF02801; ketoacyl-synt C; 6.
 DR Pfam; PF00550; dp-binding; 6.
 DR TIGRFAMs; TIGR00128; fabb; 6.
 DR PROSITE; PS00075; ACP_DOMAIN; 6.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
 DR PROSITE; PS00112; PHOSPHOPANTETHEINE; 5.
 KW phosphopantetheine; Transferase.
 QY SEQUENCE 11096 AA; 1150415 MW; 776CAEAFCAE551DD CRC64;
 Query Match 30.5%; Score 236; DB 2; Length 11096;
 Best Local Similarity 1.1%; Pred. No. 51;
 Matches 99; Conservative 18; Mismatches 31; Indels 8882; Gaps 30;
 QY 1 AAG----- 3
 Db 384 AAGVTKMIMALQGVLPRLSHATEFTTDDWTAGSVLLDETVAWPGTGRARRAGVSFG 443
 QY 4 ----- 3
 Db 444 LSGTNAHVILEQAPTAPEEPTTEPTVRPAVVPWALSARTAAALDAQRLTGHLDTPDA 503
 QY 4 ----- 3
 Db 504 DPLDVGVALADGRATFEHRAVLLPDGTELAHGTAGECAVLFSGQGSQRPGMGRHLAR 563
 QY 4 ----- 3
 Db 564 FPFVAAAFDEITALLDTHLDRPLREVVGWTDADLLNDGTGWAQPALFAVEALYRLVASLG 623
 QY 4 -----GILHLE----- 9
 Db 624 VTPDFVGHSIGELAAAHVAGVLSLEDACTLVAARARLMQALPRGAMLAIRATEDEVTP 683
 QY 10 ----- 9
 Db 684 HLTDDVSIAAVNGPTSVVVGTEEAIAIGARFTAQDRKTRRLRVSHAFHSPMDPMLAE 743
 QY 10 ----- 9
 Db 744 FRAVAAGLTYHEPRIPVLSNLTGTVAADVADLCSADYVVRHVREAVRFADGTALTDRGVT 803
 QY 10 LLVAVGPD----- 17
 Db 804 TLVELGPDGVLSAMAQESLPDGMAAVPLLRKDRPEELSAVTGLARAHVGVTVRWAGLFD 863
 QY 18 ----- 17
 Db 864 GTGARRADLPTVPFOHQRFWPTAARAADVTAAGLGAADHPLLGATVELADGAGYLFTSR 923
 QY 18 ----- 17
 Db 924 LSVRTHPMLADHGVQGRALLPGTAFVELAVRAGDEAGCDRVEELTLAAPLVLPERGQVOL 983
 QY 18 ----- 17
 Db 984 QVRVGAPDAAGRRITLIGFSRVEDGFDLPMSQHATGVLTAGAGAPDPTFDATVWPPSGARP 1043
 QY 18 -----VFO-----AHQEDTERY----- 29
 Db 1044 VDLTGAYERLAALGFOYGFAPGGLRAAWRRDTEVYAEVALPDGADTDPAAFGHLPALLDA 1103
 QY 30 ----- 29
 Db 1104 AQHAAAYADLGAISRGGLPFAMEGVSLAAAGATTVRARIAPAGEDTIVTIAYDAAGTVL 1163
 QY 30 ----- 29
 Db 1164 SVDSLVSREVVPADAPGAACTVHRDSLHVEMTPLOGRPGPAPATVAVLGPDPDALADTLR 1223
 QY 30 -----VLTNL----- 34
 Db 1224 ATGIRTTAPRDLAALADAEQVVDLVVTTLTTPGAPVPDAAHAHTAAVLAQAQWLADD 1283

QY 35 ----- 34
Db 1284 RFADARLVLRGATDGTDPAAAAAGGLIRTARTENPGRFALLDLAPDTRPDPETLAYA 1343
QY 35 ----- 42
Db 1344 LAASHDEPDIAVRGTDVHAARLARVPLATEPTTWNPDGTVLITGGTGGGLGAVLARHLVAT 1403
QY 43 ----- 45
Db 1404 HGVHLLLASRRGPAADGADDLTAELTGIGATVHIAACDVADPAALADLLGTVPAGHPLT 1463
QY 46 ----- 45
Db 1464 VVHTAGVDDGVLSLTPQRLDQVLRPKADAAWHLHEATRHLDLDAFVLFSSVAATLGS 1523
QY 46 ----- 45
Db 1524 PQQANYAAGNAFLDALAARRAATGLPATSLANGPWTOSVGMTSSLSLDLVERIARSGMPP 1583
QY 46 ----- 45
Db 1584 LTLEQGTALFDAALAAGPAALAPVRLDLPVLRTOGDIAPLLRGLIRTPVVRTAAQVSQTA 1643
QY 46 ----- 50
Db 1644 DGLAQLAGLDAAARREALLELVRTOIAQVLGHADATEVETGRQFQDLGFDLSLTAVELRN 1703
QY 51 ----- 50
Db 1704 ALNTATGLRLPATWVFDYPTPHALADHLRDELLGTEAESTAVPVRTRTAGTDPIVING 1763
QY 51 ----- 53
Db 1764 MACRYPGIASPEDLWRLVSGADATGFFPTNRGMWLDNLVDPPDRPGRTHVRAGGFLH 1823
QY 54 ----- 53
Db 1824 DAGSFDADFQMSPREAMATDSOORLLELSWEAVERAGIDPASLRDSGTGVFAGVWYND 1883
QY 54 ----- 53
Db 1884 YGTTLTGDEYEAFRNGSAPSVASGRVSYTLGLEGPATVDTACSSSLVALHWAQAALRA 1943
QY 54 ----- 53
Db 1944 GECSLALAGGVTVMSTPSTFVEFSRQRLAPDGRSKAPAEAADGVANSEGVMGLVLERQS 2003
QY 54 ----- 53
Db 2004 DAVENGHEILAVVRGSAVNODGASNGLTPNGPSQORVIRQALASGGLSTADVDVAEAGH 2063
QY 54 ----- 58
Db 2064 TGTTLGDPIDIAQALLATYGRDRDENPPLLGSIKSNIGHTQAAAGVAGVIKWMMRHGV 2123
QY 59 ----- 62
Db 2124 LPOTLHVDAPSSHVDWSVGAVELLTEQTAWPETGRARRAGVSSFGISGTNAHVIEQSPT 2183
QY 63 ----- 62
Db 2184 AVPATPASADRSVEEPPAVPWALSGKTPDALRDQAARLLAHVEAHPALRPVDISYLIAT 2243
QY 63 ----- 62
Db 2244 RTAFDRAVAVLGTDRAEALRALTAAGETDPAALTGTVTRGTAFLEFSGOGSORLGMGR 2303
QY 63 ----- 62
Db 2304 VLYERFPAPAEALDVTALDAELGHPLURDIWGEDAQLVDRGTGTQPALFAIEVALFRL 2363

QY 63 ----- 62
Db 2364 LEAWGITPDFVAGHSIGIEIAAAHVAGVLSLGDACELVVARAVLMOSLPEGGAMIAVQATE 2423
QY 63 ----- 62
Db 2424 DEVLPLLTDDVSIAAVNSPTSVVSGYENATLAVARHFDQGRRTTRLRVSHPHSPLMA 2483
QY 63 ----- 62
Db 2484 PMLDDFRAVVESLJTTFTPTTPVVSNLTGELAPAEALCSADYVWRHVREAVRFADGIRTLA 2543
QY 63 ----- 69
Db 2544 DRGVTTFVELGPDPSVLSMAQESAPGAGTIPLLRRDRPEEQAVLAALCHLQVLGVEADM 2603
QY 70 ----- 69
Db 2604 SATFRGLDPVRVDLPTYAFQHRWFMPAARPARPDDVRAAGLGAAEHPLLLGAAVQLPDDDG 2663
QY 70 ----- 69
Db 2664 ALFTGRLSLRTHPWLADHTVLGTVLLPGTALVELAVRAGDETSCHLEELTLAAPTUPE 2723
QY 70 ----- 69
Db 2724 DGATLLQVRVGSADDTGRTVTVHARPDDTADRTWTLHATGVLATTPPAAAAFTTVWPP 2783
QY 70 ----- 69
Db 2784 ADAEPLTTDDCVAHFTTHRFAYGPAFOGLRAAWRAGDVLVYAEVALPESATDEAAAFGLHP 2843
QY 70 ----- 69
Db 2844 ALLDAGLHAALLADDDRTGLPFSWEGVTLHASGATALRVRLAPNGPNSLVTAAADPAGNP 2903
QY 70 ----- 72
Db 2904 VATVTRLLARPDLASQLTIHSALTTRDALPHLDWTVPPLPDTANSAPPALLGPDATVLADA 2963
QY 73 ----- 80
Db 2964 LGDPAVARHATLDDLGLAGDTTPPTVPLGAPLDGDTAQHAHALTRLSALTILVQOHWLATD 3023
QY 81 ----- 80
Db 3024 RLADSRLEFVTHGAVATDDAPPTDLAAAAMVGLIRSAQENPGTFTLLDLDEPDSTTAL 3083
QY 81 ----- 80
Db 3084 SRALTILDEPQLLLRAGRARAARLTRPAPTTTTHTPWSADGTVLVTGGTGGGLGVARHL 3143
QY 81 ----- 82
Db 3144 VRSCGVHLLLTSSRGVGAAGAAGLVAELESIGARVVAACDVGDGSAVAELVAGVSESY 3203
QY 83 ----- 82
Db 3204 PLSAVVHAAAGVLDDGVVGSGLTPTERLAALVRPKVDGAMNLHEATRGDLDDAFVVFSSVAGV 3263
QY 83 ----- 87
Db 3264 FGAGQANYAAGNAFLDALMWHVRVAGGLPGVSLANGAMWOGVMTAGLTERDVRRAESG 3323
QY 88 ----- 87
Db 3324 MPLLTVDOGVALFDAALATGSAALVPVRLDLAALTRGRDIAPLLRLGLVRAPLRRRTAATGL 3383
QY 88 ----- 87
Db 3384 ATGADTGLVORGLRDLPHORHEALLDMVRSSAALVLGHADGNAIDAERAFLRDLGFDLSLTA 3443
QY 88 ----- 87

Db 3444 VELNRRLRTATGLHLSATWVFDHPTLSALAEHLRDELFGAVESEVRVPVQALPPTADDP1 3503
Qy 88 -----NPED-----DTPDGH----- 97
Db 3504 VVVGMACRFPGVTSPEDLWPLVDGCTDAITTFPTNRGWLDNLYDPDPHFPGTSYTRSG 3563
Qy 98 -----:||||----- 97
Db 3564 GFLHEGEFDPAPFGMPREALATDSQORLLESSEWAEIERAGIDPLTLRGSATGVFAGV 3623
Qy 98 -----:||||----- 97
Db 3624 MYSYGSILGKEGEGGQGSAGSVASGRVSVYALGEGPAVTVDTACSSSLVALHWAQ 3683
Qy 98 -----:||||----- 97
Db 3684 ALRAGECSLALAGGVTVMSPTSFVEFSRORGLAPDGRSKAFAEADGCVGSEGVILVL 3743
Qy 98 -----:||||----- 97
Db 3744 EROSDAVRNGHEILAVIRGSANQDASNLTPNGPSQORVIRQALASGLSTADVAV 3803
Qy 98 -----:||||----- 97
Db 3804 EAHGTGTTLGDPIEAQALLATYGRDRDPENPLWLGLSKSNIGHTQAAAGVAGVIKWMAM 3863
Qy 98 -----:||||----- 97
Db 3864 RHGVLPTLHVDPSSHVDWSVCAVELLTEQTAWPETGRVRRAGVSSFGISGTNAHVIVE 3923
Qy 98 -----:||||----- 97
Db 3924 QPALVESPAAPESGREGVVPPLSGKSPALRDQAARLGLAERPALRPLDLGYSLAT 3983
Qy 98 -----ADL----- 100
Db 3984 TRSAFHRVAVLATRADAVRALTAALAAADADLSAVGVGTRTGRHAVLFSGGQSORLGMG 4043
Qy 101 -----:||||----- 100
Db 4044 RELYERFPVFAEALDVAIDHLDAALPAQASLREVVMGDDVELLDEGTGTQPALFAVEVAL 4103
Qy 101 -----:||||----- 100
Db 4104 FRLVESGVRPDFVAGHSIGEIAAAHVGVFSLEDACRLVAARATLMQALPTGGAMIAIQ 4163
Qy 101 -----:||||----- 100
Db 4164 AAEDEVTOHLTDDVSIANAANGPSTVVVSGAESAAARTVADRLEANGRKTTTLRVSHAFHSP 4223
Qy 101 -----:||||----- 100
Db 4224 LMDPMLAEPRAVEGLSYATPTLPVVSNLTLRLATADDLCSAEYWARHVREAVRFADGVS 4283
Qy 101 -----:||||----- 100
Db 4284 TLNEGVTTFLELPGDGLVLSAMAQOQLTGDAATVPALRKDRDETSALTALAHLHTAGLR 4343
Qy 101 -----:||||----- 100
Db 4344 VDMAAFFAGSGATRVLDLPTAFAQHATYWPVTGTLPTAHAAAVGLTAAEHPLLNGSVLEAEG 4403
Qy 101 -----:||||----- 100
Db 4404 EGVLFTRLSLQSHWPLADHAVMGVQLLPGTALLELAFRAGDEACDRVEELTLAAPLVL 4463
Qy 101 -----:||||----- 100
Db 4464 PERGAVOTQVRGVVADDTGRRTVTVHSRPEHATDVSWQHATGTLTMSAPADTGFDATA 4523
Qy 101 -----VLY----- 103
|||

Db 4524 WPPADAEPLATDDCYAREFTTLGFAYGPVQGLRAAHRAGDVLVYAEVALAESTGDEATAPG 4583
Qy 104 -----:||||----- 103
Db 4584 LHPALLDAALHASLVAHEGESNGGLPFSWEGATLYATGATALRVRLTPTGTGDRSVAIA 4643
Qy 104 -----:||||----- 103
Db 4644 VADTAGRPVAAIDNLVSRVSGDQLTGAAGLARDALFTLDWNPVNPENLVNPVPENTGG 4703
Qy 104 -----:||||----- 103
Db 4704 CHAQDQGRPAAATVALVGADGTAIAADLTAAGIHTTLHPOLTTLATTADVPKTVLPL 4763
Qy 104 -----:||||----- 103
Db 4764 TGTGTGTGTGTGIGTGAAESDASAPSAEVAHTLSTAALALVOEWTAOERFAGSRL 4823
Qy 104 -----ITRPD----- 108
Db 4824 APVTTGATAAGTDVMDVAAAVMGLVRSASOEA PDTFVLIDRDPGPAGTHDRTAAABERG 4883
Qy 104 -----:||||----- 108
Db 4884 QLLRLALHTDEPOLALRDGVLAAARLAFDTAAALTTPADRAWRLDSTAKGSLNGLALTP 4943
Qy 109 -----:||||----- 108
Db 4944 YPALALPLTGHEVRVEVRAAGLNFRDVLNMGVPGDDVSGSFAAGVVEVVEPVTGL 5003
Qy 109 -----:||||----- 108
Db 5004 APGQVMGMITGSGSLAVDDARLRLPEDMSWETCASVPLVFLTAYYALKELGGLRAG 5063
Qy 109 -----:||||----- 108
Db 5064 EKVLVHAGAGGVGMAAIOIARHVGAEVFATASEGKWDVRLSLGVADDHIIASSRTLDFEAA 5123
Qy 109 -----LEL----- 111
Db 5124 FAEVAGDRGLDVVLSLAGDFVDASMRLLGDDGRFLEMGKTDTRAADSVDPGLSYQSFDL 5183
Qy 112 -----:||||----- 111
Db 5184 AWVPETIGTMLAELMDLFRGTALRPLPVRTWDVRHAKDAFRFMSMAKHGKIIVLTLP RS 5243
Qy 112 -----:||||----- 111
Db 5244 WKPEGTVLVGTGGTGLGLVARHLVRSQVVRHLLLTLSRSGVGAAGAAGLVAELESGLARV 5303
Qy 112 -----:||||----- 111
Db 5304 VVAACDVGDGSAVAELVAGVSESYPLSAVVHAAGVLDGVDGVLTPERLAALVRPKVDGA 5363
Qy 112 -----:||||----- 111
Db 5364 WNLHEATRGDLDAFVVFSSVAGVFGAGQANYAAGNAPLDALMVHRVAGGLPGVSLAWG 5423
Qy 112 -----:||||----- 111
Db 5424 ANDQGVMTAGLTERDVRRAAESGMPLLTVDQGVALFDAALATGSAAALVPVRDLAALRT 5483
Qy 112 -----:||||----- 111
Db 5484 RGDIAPLLRLGLVKAPIRRAAATTPGDTGLAEQLTRLQRAERRDTLLALVRDQAAVMVLGHT 5543
Qy 112 -----:||||----- 111
Db 5544 SGDGVDPRAFRDLGDFSLTAVELNRNIGAAITGLRLPATAVFDYPTADALAAHLLTELLG 5603
Qy 112 -----:||||----- 111
Db 5604 PDAESDPDEPGDPTAGTDDPIVITGMSCRFPDIGSPEDLMRLLDGADVVDTDFPTNRG 5663

Qy	112	-----	111	Qy	125	-----	124
Db	5664	WDLNLYDPDPAHAGTSYARTGGFLHDAADFADFCGMSPREAMATDSQORLLLESSWEA	5723	Db	6744	PDGEPTRTGVPFWSRGVFLAASGASSVRVRLGRDSDGTLAIADTTGAPVASVOALSM	6803
Qy	112	-----	111	Qy	125	-----	124
Db	5724	IERAGIDPLTRDSRTGCFAGVMYSGYGTRLDGAIEFGOGSALSASGRVSYTFGEE	5783	Db	6804	RTVSVTALSATAGLARDALFRLDWSAPEPACQPDVTVTIPIPAVAVGTSETSELTA	6863
Qy	112	-----	PDGN 115	Qy	125	-----	124
Db	5784	GPAMTDTACSSSLVALHAAQALRGCECTLALAGGVTVMSIPDTFIEFSRQGLAPDGR	5843	Db	6864	ALRAAGADVVRTTLLSTDEPAPALIALPLVASDQTGTAEAAAPVPAVHDLTRRALALVQT	6923
Qy	116	-----	115	Qy	125	-----	124
Db	5844	SKPFESADGVGMSGVGMLLERQSDAVRNHQILAVVRGSAVNODGASNGLTAPNGPS	5903	Db	6924	RLQEOHFADTKFVFVTRGATVGRDVAAAAMGLVRSQAQENPCGFALVDLDPDGAVGAAA	6983
Qy	116	-----	115	Qy	125	-----	124
Db	5904	QORVIQALASGGISTADVADEAHGTGTTLGDPIEAQALLATYGRDRDPENPLILGSIK	5963	Db	6984	LVAALVSGEPQLAVRGDVLRLVARLVRRPLTEVGACADGTGCGVGGSGVSPSGEGAVLVT	7043
Qy	116	-----	115	Qy	125	-----	124
Db	5964	SNLGHQAAAAGVAGVIKWMMRHGVLPRLNITEPSSHVDWSAGAVELLTEQTAMPETG	6023	Db	7044	GCTGGLGAVLARHLVAEYGVVRDLLLVSRGERAVGAGELVAELAGVARGVRVACDVTR	7103
Qy	116	-----	115	Qy	125	-----	124
Db	6024	RARRAGISSFCISGTNAHVILEQPEAAARHSAPEEADTAAAKAPATAHLVMPWALSGK	6083	Db	7104	AAVVELVGHAVSAVVHAAGVLDGCMVGALTGERLSAVLRPKVDAMVHLHEATRGDLDA	7163
Qy	116	-----	115	Qy	125	-----	124
Db	6084	TPEALRAQARLLAHLQORPELAPADIALSIATORSOFTHRVVLTDRDEATRLSALA	6143	Db	7164	FVVSSLAGVFGSGOANYAANAFLDALMTRRRRAEGLPGLSLANGPMSLTDTGTCMLAD	7223
Qy	116	-----	115	Qy	125	-----	124
Db	6144	TTAASDPALTGTVTMGRCAVLFSQGSORGLMGRELYERFPVFAEALDVVIDHDAALP	6203	Db	7224	ABADRLTRSGVPLTAEQGLALFADALATGATCVPRLDLSALRAQGEVPPLLRSLIRG	7283
Qy	116	-----	115	Qy	125	-----	124
Db	6204	AOAGLREVMGDDVELLNETGTOPALFAIEVALFRLVESWGVRPDFVAGHSIGEIAAAH	6263	Db	7284	RSRRAAAESATATGLRLRVLGLNPVERQEVLLDLVRGOVALVLGHADADDVHPARAFRE	7343
Qy	116	-----	115	Qy	125	-----	124
Db	6264	VGVFSLBEDACRLVAARATLMOALPAGGAMTAVQATEDEVIPHLTDEVAIAVNGPISVV	6323	Db	7344	LGFDLSITSVELNRNLNTVTGLRLPATMVPDYPTVEVLVSIVLDELGTDAEVATVQPAAV	7403
Qy	116	-----	115	Qy	125	-----	130
Db	6324	ISGAEEATQVAQHPADQGRRTTALRVSHAFHSPLMLAEFRAVAEGLSYATPTLPVVS	6383	Qy	125	-----	130
Qy	116	-----	124	Db	7404	AVADDPIVIVGMACRYPGGVASPDLLRLVTDCGVDAVSPPTNRGWDVESLYHPDPHLG	7463
Db	6384	LTQVATADELCSAEYVVRHVREAVRFADGVTALAEAGVTFLELGPDPGVLAAMARETVA	6443	Qy	131	-----	130
Qy	125	-----	124	Db	7464	TSYTRSGFLHEAGEPDFGFGMSPREALATDSQORLLLESSWEAIERAGIDPVLSRGR	7523
Db	6444	DDTVTVPLRRNMPEERTLLTALGRHTTGPIDWAALLAFTGARPVDLPTIYAFQHRFPW	6503	Qy	131	-----	130
Qy	125	-----	124	Db	7524	TGVFAGVMYSYDYSAMLASPEFEFGSGSSPSLASORVAYTLCLEGPVAVTDTACSSSLV	7583
Db	6504	PSGPRDTADAAVGIAGASHPLLNGIVELADEGLLFTGRLSLQSHPLADHAVMGQVLL	6563	Qy	131	-----	130
Qy	125	-----	124	Db	7584	AMHWAMQALRSCEGLALAGGVTVMTSTPAVFDFAQRGLSPDGRCKAFADAAGVWSE	7643
Db	6564	PGTALLELALRAGVEGDHVEELTAAFLVLPERGAQVQTVRVGVADTTGRTVTIHSR	6623	Qy	131	-----	130
Qy	125	-----	124	Db	7644	GVGVVLERQSDAVRNHGHEILAVVRGSAVNODGASNGLTAPNGPSOORVIRQALASGGLT	7703
Db	6624	PARATTTSDTHGTDTPTWQHTAGVLVAGLPATATVPFDATVPWPAHAEPVDLADFYAS	6683	Qy	131	-----	130
Qy	125	-----	124	Db	7704	AGDVVVEAHGTGTTLGDPIEAQALLATYGRDRPERPLLLGSVKSNLGHTQAAAGVAGV	7763
Db	6684	RAGEFGYGAFOGLRAAMRRDGEVFDVALPEAGRTEAEAYGLHPALLDAGLHAALWA	6743	Qy	131	-----	130
Qy	125	-----	124	Db	7764	IKQVLMRHWGVVPRTLHVDAPSSHVDHSEGAVELLSEQAAMPETGRVRRAGVSSFGISGT	7823
Db	6744	-----	6743	Qy	131	-----	130

Db 7824 NAHVILPEAARRPVMTNETVPSTVPWVLSGKTPALRAQAALLSSIEERPELRVD 7883
Qy 131 ----- 130
Db 7884 VGMSLVTRSTFEHRAVLAAADRAAARALSAIADEADAAAATGRVGRHVLPSGOG 7943
Qy 131 ----- 130
Db 7944 AQLGNGRELYERFPFAEALDVVVDHDLAALPAQAGLRVVMGDDAELLNETGTQPAL 8003
Qy 131 ----- 130
Db 8004 FATEVALFRLVESGVRPDFVAGHSIGEIAAAHVAGVFSLEDACRLVAARATLMQALPAG 8063
Qy 131 ----- 130
Db 8064 GAMIAOATEDEVPHLTDVATAAINGPNALVVSVEDAAVEIGARFAAEGRRTRLHV 8123
Qy 131 ----- 130
Db 8124 SHAFSPLMDPLAEFRVVAEGLSYAAPSLLPVVSNLTGQVATADELCSAEYWRHVREAV 8183
Qy 131 ----- 130
Db 8184 RFADGVTALEAEGVRTFLELPGDGVLAAMAGASLTSSLAVPLLRKDRPEEPAALAAQ 8243
Qy 131 ----- 130
Db 8244 LHIAGARVDPVLPFAGVGAGRVELPTYAFORGWFWPVRGVGVDGAVGLSAGHPLLG 8303
Qy 131 ----- 130
Db 8304 AVELAAGAGVLTGRLSLSSHGLADHAVMGRVFPVGTALLEVMVRACDEVCGRVEEL 8363
Qy 131 -----TWSC----- 134
Db 8364 TLAAPLVLPERGVRVQVAVDAPDAAGRRGVGVYSCPDGVGQAVMSHAGVLSAGVADQ 8423
Qy 135 ----- 134
Db 8424 VGGFGDGVWPPQOGAVSDAEGCYELPADAGFGYVFOGLRAVWRGGEELFAEVALSDE 8483
Qy 135 ----- 134
Db 8484 VAESADTATGFLHPALLDASLHASLLSLEGOSADGCPALPFAWEGVSLFASGATALRV 8543
Qy 135 ----- 134
Db 8544 RLAPAGEHAVSVTAVDPTGAPVISIDALTRRLTLDEVNASHTQSLDALFGVQWTTVPST 8603
Qy 135 ----- 134
Db 8604 PAADHPSVAIIGTDHLGLAEALSSSAGATTTTAAAYESLDALIAAGPEVSPDVTLIG 8663
Qy 135 LITEDT----- 140
Db 8664 LTTEDAIAQVYNDHDTAVAGGTIGACAAAADAARLTAELRTIQAWLADERLAARLV 8723
Qy 141 ----- 140
Db 8724 FVTRGAADGQDVAAGVGLVRSQAOTENPGTFLGLDLDTGTEASTAVLGEALTSDEPOLLL 8783
Qy 141 ----- 140
Db 8784 RDGHLHAARLRLASPADTAVPTENNADGTVLITGGTGLGAQPARHLVDRYGVNRLLLV 8843
Qy 141 ----- 140
Db 8844 SRRGPDPAGTTTELVAELTAHGAEVAQVQADGDVAALVAGVDEHPLRAVHTAGVL 8903
Qy 141 ----- 140

Db 8904 DOGVIGSLTEERLATVLRPKADAAWHLHEATRGDLDPVVFSSVAGVFGGAGONYAAA 8963
Qy 141 ----- 140
Db 8964 NAFLDALMAQRRAAAGLPGLSLAWGPDQGTGWTGMLSDAEADRLARSIGPPLSABOGLAL 9023
Qy 141 ----- 140
Db 9024 FDAALALAGTSTPDRAAGSAAASTGTGTITAIIPAAALVAPVRLDAAALAAQGEVPAILR 9083
Qy 141 ----- 140
Db 9084 GLVTRTRTRTAGGSVTVAGLVNRLSGLTADERRQELLELVRTQAAVLGHADPASVDST 9143
Qy 141 ----- 140
Db 9144 AQFRDLGFDLSLTAVELNRRLSTATGLRLTATLVFDYPNTDALAEHLRDELFGAVESEVRV 9203
Qy 141 -----GFDL----- 144
Db 9204 PVQALPTADDPVVVGMACRPPGGVTSPEDLWRLVDAGTDAITFTPTNRGWDLESYDP 9263
Qy 145 ----- 144
Db 9264 DPAHLGTSYTRSGGFLHEAGEFDPAFFGMSPREALATDSQORLLLESWEAIERAGIDPL 9323
Qy 145 ----- 144
Db 9324 TLRGSATGVFAGVMYSDYGSILGGKEFGFGQGSAGSVASGRVSYTLGFEPAVTVDTA 9383
Qy 145 -----GVTI 148
Db 9384 CSSSLVALHAAQAALRAGECTLALAGGTV 9413
Qy 145 ----- 144
RESULT 14
Q18559 PRELIMINARY; PRT; 7829 AA.
AC Q18559;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Hypothetical 874.1 kDa protein.
GS C41A3.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX Bentley D.;
RT "The sequence of C. elegans cosmid C41A3.";
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U41541; AAK18894.1; -.
DR HSSP; P25715; 1MLA.
DR InterPro; IPR001327; AC transferase.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatr.

DR	InterPro: IPR000794; Ketoacyl-synt.
DR	InterPro: IPR000734; Lipase.
DR	InterPro: IPR003880; Ppantne_attach.
DR	InterPro: IPR000379; Ser_estrs_site.
DR	Pfam: PF006698; Acyl_transf_2.
DR	Pfam: PF00501; AMP-binding_2.
DR	Pfam: PF00668; Condensation; 1.
DR	Pfam: PF00109; ketoacyl-synt; 5.
DR	Pfam: PF02801; ketoacyl-synt_C; 5.
DR	Pfam: PF00550; pp-binding; 8.
DR	PROSITE: PS00075; BcP_DOMAIN; 8.
DR	PROSITE: PS00606; B_KETOACYL_SYNTHASE; 4.
DR	PROSITE: PS00120; LIPASE_SER_UNKNOWN_1.
DR	PROSITE: PS00013; PHOSPHOPANTHETINE; UNKNOWN_3.
KW	Hypothetical protein; Phosphopantetheine; Transferase.
SEQ	SEQUENCE 7829 AA; 874065 MW; E19BDDA392F325C1 CRC64;
Query Match 30.1%; Score 233; DB 5; Length 7829;	
Best Local Similarity 1.3%; Pred. No. 25;	
Matches 90; Conservative 25; Mismatches 33; Indels 6699; Gaps 27;	
Qy	1 AAGGI----- 5
Db	151 AAVGIETACSSSLVAFHLARQAIQSGETKLAVCGANHVGRSFHLYNSHMVSPNGRLA 210
Qy	6 ----- 5
Db	211 AFDRSANGFVRAESFAVLCSKOPAEENLLIHCEVCVSFNSDGKTPSLTAPNPISQY 270
Qy	6 ----- 5
Db	271 EVQLEALKNIKDQSVQLVTCHTGKGLDQVELTAINRSFKSDIRVMSPKSSMGHGCAA 330
Qy	6 -----LHLEL----- 10
Db	331 GLIGVLQSLYSMHGIIIPNLHLELPSDLGDKSMGFVMEELNRAISSYGFSGTNA 390
Qy	11 ----- 10
Db	391 CAITEKPKSPVLQKESVAESNVFLSAKSHESLKLQIEEYTFQFMAQSDAMEDILYTVN 450
Qy	11 ----- 10
Db	451 ERKTYDFRAAVFGKDNSEIARKLDQGDYSLTNLQESTFEVEFGEGNEKLLRLMYEKN 510
Qy	11 ----- 10
Db	511 ETFHSTVDKYCKLAETCGPPEARALTAFPFPKLTPLTYNVRSLISSMATFELLVOYNLT 570
Qy	11 ----- 10
Db	571 PNKLRGKGLGQIFCLAVAKVITFESAQLIKGVAAEANTLDILGIELKSSKPIEIQHL 630
Qy	11 ----- 10
Db	631 KSTKKKILPIHISGELKETAKPNLMTFIVNGEILLELDPVRKVQKLIQLFACGDFPAVK 690
Qy	11 ----- 10
Db	691 PRGRIVKTPYSLKRFQFWPEVOTAMTNIVDEQTNSSLSDAEISTVRTIVKQFLDIEE 750
Qy	11 ----- 10
Db	751 DDINLETCAVDSLTSIEMVEAFGTAVNOTMPFDLLLEAYPTILNIVDFLKLTVTVPTVK 810
Qy	11 ----- 10
Db	811 ATTSIHKTKSELSTSDINVIADYQFAGVEGKELWDTLLTSRLTTGKISDIRKKQCEGD 870
Qy	11 ----- 10
Db	871 AGLEVGLLKODTSMFONSFFAIAKBAEFLDPQHRLLLNAAYNALSKSLTSPDADLFL 930

```
Db 2011 QDMMESSILAMLSMAAIVAGRVYIFGCGPSVTIDTACSSSLVALEMAINALLDNRC 2070
QY 17 ----- 16
Db 2071 SKVIVAGVNLILNEKQGLRTNGKMLSQHMSLSFDSRSGYGRSDCCVVLMELEAKPNF 2130
QY 17 ----- 22
Db 2131 HYMTTQSVNVNHGGRSVSLTAPNGVAHKMLLTSTVINQSPSLAIDYWEAHGTGTPLGDPI 2190
QY 23 ----- 22
Db 2191 EFWTLSSILONIIIGSVKASLGHEASAGTCGLLKLFLMLTYQVPTLIHFHVLNKDINA 2250
QY 23 ----- 22
Db 2251 GSIRLPIIGEDSELVSAGISSFGVSGTNAAAIAFNNDNNKLEPIPIHKYIILPISAKNOI 2310
QY 23 ----- 22
Db 2311 SLDNLEKQILSVIPLTDVPTCNIASALANNRSHFTIRNALIVNSGIVNSKMEGKPHRVA 2370
QY 23 ----- 22
Db 2371 KDRYHVHKLCDLLDASLLQYDVINETYTVASLKNPQSPAMKFAIIFELTSLEYEIVA 2430
QY 23 ----- 22
Db 2431 SDGEELLAVLLANGSLKWNFNKMTIELPIGSLLTETAFADHDLNSTTSSSIKSYQTQPESH 2490
QY 23 ----- 22
Db 2491 NLDSPMELMKLIMKLYITGYDVWATVYSPVEQFIALPNYQFNKQTLWFEERLEIVDHYL 2550
QY 23 ----- 22
Db 2551 IGTIDESEDTLLKNOISELRHPQFKGKPLDVGTMSEITAIEALKIRNEIPEFSIQNLKT 2610
QY 23 ----- 25
Db 2611 ELITLTKPAMLETNVTNRNEDEGFNVSAIDGQRLFSLNASSVEIONIEVPAVEVQIPDK 2670
QY 26 ----- 25
Db 2671 VVYLKECPNAVIRRRHNMVYVDSRAEQSPRTANIVLNEIIGFAPTPSDMFIILGVLP 2730
QY 26 ----- 25
Db 2731 VHYMVQYDGDALMQFQMSODKRVLSNIYVLKDAKGLEIPTIRMHKKSTLLSSQEASIVA 2790
QY 26 ----- 25
Db 2791 AKTLQMAVRHKVCLAVGDVIESGLDIDESQSTGFSGLDVLATVLLNRLNQYFPEI 2850
QY 26 ----- 25
Db 2851 ELTTSDLPDNPISIIDLSIMIEQLLNEKGITEPSEPNTPKTSLRGRKLSIPAVRAQVLAQI 2910
QY 26 ----- 25
Db 2911 EFVENYNSKQKQVQAEAPSSSSSECSNHLBESDATVDRTEIRRKVSLAVFDLATETLSAED 2970
QY 26 ----- 25
Db 2971 LQSKGTGLGMSLSIVDFVNLNDKYFPDDITASDIFDYPTVDELSDHIVRKSSSVP 3030
QY 26 ----- 25
Db 3031 PAASEIMKETWNGISTVDAEHTKLENLSOSFMLENQNSINPTLKMWSNQITKLVKPS 3090
QY 26 ----- 25

Db 3091 DGNFLFELNANGOEKEIQHFTGPNNIIIDLKGFHEGSTETLYMSLLNLVKSISKLEIQ 3150
QY 26 ----- 25
Db 3151 CRFGVSQFGLGNSISRPFMTVAAEKNPLISFPAWYQNVQVSPVDSPTGNWLTGG 3210
QY 26 ----- 25
Db 3211 LSGIGLEIGKFIANGAENVILISRRQPTAKALRDLSTELTHIGLARKTIVLKLKNIS 3270
QY 26 ----- 30
Db 3271 AKLIVFQSKLSFSTKVKYISDLFSKKVTFFYNLLQSKCFSKHFHWFLEVLVIKQWTV 3330
QY 31 ----- 43
Db 3331 HTIAADINDKEKLRELTKLVNGITGIHSAGVLKDSKIERQNKESPNQVPTPKANGPHV 3390
QY 44 ----- 43
Db 3391 LEEIEKHFNKYIENFIMSSFTAACNEGQLNYGVSNAYLEYQVORRRRQCKSGCAIOWG 3450
QY 44 ----- 43
Db 3451 NWIDTGMATDENVRKFLANLQFLOGHNKDALKYLACILTRPBLIMVANIDNVLKNRK 3510
QY 44 ----- 43
Db 3511 DLPKDLINTGILPPEDFTGKNNESEFPPLSGNDFEKVSNFSEVEDEEVLLEKEKVSIL 3570
QY 44 ----- 43
Db 3571 MCSPTKLKNNKINMDGLDLSKLIVEFLNFINSTFKISVNLSDAYNHPHLEKLAHI FQOM 3630
QY 44 ----- 43
Db 3631 TIVDHPVNSVKEEIEFKSTDFCPIFGINIFPNKNDPDKAKSTAVKLLENGELTAGKY 3690
QY 44 ----- 49
Db 3691 AVSVVGKSTRDVVSKIKEAAPQIKLCQESSKVLMTQSGSOYPMMGROLVENYEIPT 3750
QY 50 ----- 49
Db 3751 TLQSLKKCDEYLOGDVSLEILFNTDHYKLLQLTQHMOPIMFCFGYATAQLWLSLGI 3810
QY 50 ----- 50
Db 3811 DYYLGHSVGLVAGVLAGIMSIEDGLRLIVERGKAMENIAGLCALLAVOREIADEVLRKF 3870
QY 51 RV-----HLVKM-- 57
Db 3871 KVSATINSPQVVFAGTKSVLDAALAFVKGQKQATYVNOQYPFHSNLIQETHLVSLRQ 3930
QY 58 ----- 57
Db 3931 CLADIKFSAGRTPLVSNVTGQIINTFSEAYIVKHTVSAVKFVDCVETLOAKGVTVWIDAG 3990
QY 58 ----- 57
Db 3991 SAAVLATFVKRIIPTELSKHRIVOTCKEKESDVNLVOACLELEQSGLPISWTLTGCG 4050
QY 58 ----- 57
Db 4051 RNADERLVEFPMTHNDIIKNDEPHELLEGHQLNGKIWVAGAYQFKIDQLVKLKAAGWELM 4110
QY 58 ----- 57
Db 4111 LKNVKELKPWYIEDNREYQIOWNSDMTIELIVNSVIVCSLEVEPONSVLKLETISENEKP 4170
QY 58 ----- 57
Db 4171 FEVHDFYETLFRNGLOYDSGFRRIESARRSKRCFSQIKSSPFAWPLIDSAMHSITASVV 4230
```

Qy 58 ----- 57
Db 4231 PRPDYFLPVAMGSVTMKTNSFTLPNLHAQTVITSETDKFIOVNALLAGDTPICEVR 4290
Qy 58 ----- 63
Db 4291 NMTIVLKTPEVHTRIPIINSIETVETPPKSEIEIIVGFDISLPYNIOISENSENWQHLKNTV 4350
Qy 64 ----- 63
Db 4351 KOKLHNRSLKODHARVALLDSARYDPEYFGRIPSEAKFIDPOORLLLCVAKLLDSL 4410
Qy 64 ----- 63
Db 4411 ITSLSNTGVFIGSANEFSHVAYGYKOPRAEWSGGTSALAGRIAHWLKLGPPVT 4470
Qy 64 ----- 63
Db 4471 LDTACSSSFYALSAACDALRTGQCEYAIIVGTNLVHMHMTDVLQNAKMTVDVDFCKAFDV 4530
Qy 64 ----- 71
Db 4531 DANGYKRSEAVCSMLLTSPNIDSATITNYATGHNGTSSSLFTPNGLSOLEVMQRATNP 4590
Qy 72 ----- NLTSSLS 79
Db 4591 LEKILEIOTHCTGTLGDPPIEINAIKSLVSSACKIGSVKSNIGHTEGSSGLVSLCSSLMS 4650
Qy 80 ----- 79
Db 4651 FRSKYVAQLHLKCPETNSIKTNKICRFIGEDADENNILINFGFTGSCNCSVVLKPKNA 4710
Qy 80 ----- 79
Db 4711 ISEHVSSEVFYPIILLSHSAKSLQKYVQLCEFTSNAKSLHDIMMSLFQKKIHVHRQF 4770
Qy 80 ----- 79
Db 4771 IIFNPKRIAVTSLDGFVEVRDERLEKLKPCSVFLKEGVVHFDKDKFORVDLPISVFN 4830
Qy 80 ----- VC ----- 81
Db 4831 NTLHWALSYRDEIDRHESQOMSFKNI FYEKVLTETPMQPNISKVVCIGRLDLIPKIE 4890
Qy 82 ----- 81
Db 4891 IDEVSSFANGIIVFHPISINSIFEYLKAKWMSLISRNQNVFIIICCFENGTSHTWTGL 4950
Qy 82 ----- 81
Db 4951 RSLASEKMIYPKFVSIDKVDALKLEFNHEYMFEAFIYKRGYVERLRKVNPTFKFAPQY 5010
Qy 82 ----- 81
Db 5011 EKCLISGGTGIGSAINELKPKSSVITRKNIASEDGKTFLLSSDITRLDISHKENYVYF 5070
Qy 82 ----- 81
Db 5071 LAGIVNLSLHENVKRDSDLDMVSIKLOKAKNLMKCCDETSHFVFSIANVLGSYGQSNY 5130
Qy 82 ----- GW ----- 83
Db 5131 AFSNGLVTSFLETSSTKSTIIHWGPWKVGMGLAQPERREIVKQIESNGWKLPLNODAISV 5190
Qy 84 ----- 83
Db 5191 FYTQPMETHEQIIVFDGDFDTIVARQPHLOKLLSEVWETKYKEEIKKSLNFEIPEI 5250
Qy 84 ----- 83
Db 5251 VGITDISSKINLPFMDLGIDSLCMENLRYSLNKNFDELTVMSEMPENATYQKLOTYVETL 5310

Qy 84 ----- SOTINPEDDT ----- 93
Db 5311 RKAKHSESLSHDRVSSQVSNKEDDTRVAVIGWSAEFSGSNIHEYENLMDGICSTGNK 5370
Qy 94 ----- DP ----- 95
Db 5371 YLLKNPFGDNKFNLTDARVLDPOVRKFIQHAYLALENSQYKQKHELRCGVFAGAE 5430
Qy 96 ----- 95
Db 5431 PSDYGRADDDAMRKL FVMNMNSYLAASYCLDKGBAVSVYSACSTALVAVANAVKS 5490
Qy 96 ----- 95
Db 5491 IQSGSMDYALVGAASIAEYSGALSGDFDOKTMTFSKGVCRPDKDSEIGVIRGSGVCFV 5550
Qy 96 ----- 95
Db 5551 LKRYSOALLDNNDNVHFIKIDFAINNDGHSRSMFMAPNPAGOLKCMTDVLARFTNKKEKRI 5610
Qy 96 ----- GHA ----- 98
Db 5611 SFVECHATGTTLDGTIEMNSLRTAYSPKXKLAIGSCCKANIGHAYASGLAALVKCAKMLQ 5670
Qy 99 ----- 98
Db 5671 TGIIPQVNFSEPRDGMGOFTVNGKKSTISONSLSISDSFGIGGTNVHMFIEFPARSOE 5730
Qy 99 ----- DLVLY ----- 103
Db 5731 VKKISSENILYDMIPISAKTEYSLDHTSEAI SKYLQDTNKNIAOCSSTFIHSRVPMSR 5790
Qy 104 ----- 103
Db 5791 TYLSVNNNELLKIRTNKKTWNGKSPKIALPAPOGIOFTNLPLNEYLKNSAYRREVEY 5850
Qy 104 ----- 103
Db 5851 LCEASSFGIPSLLEGILYPTKPNFDHLIHATQFAIAIFVQCMAIFKAIKNVFNPTCLIGH 5910
Qy 104 ----- 103
Db 5911 SVGEYAAVVISGALKTEALKLILKRSELIGKTEKARMLMVMNYEKOLPSHVHVSAIDA 5970
Qy 104 ----- 103
Db 5971 NTKCVGPVETIDNLEKYFINNHIKYRNITETKHGPHSKMPCISKEFEPPCESFATKVPL 6030
Qy 104 ----- 103
Db 6031 IPMISITGSEIKIFDSKYCTMHTNPVNLVVDHIMKLDIDIIIVEVGPTGVLSNLLAK 6090
Qy 104 ----- 103
Db 6091 RNSKIVVPTCGTKGHPKISLGEICIGOLWSNGVDIRKLTPLKLVADGOVPGYCFDEROFGE 6150
Qy 104 ----- 103
Db 6151 KNQKISOKLNRFSYKESKWPVSELEETKPVRTVCDGSLNGISEPVLVLLRKSEDIH 6210
Qy 104 ----- TRFDL ----- 109
Db 6211 KNYFDVELFLKSPRCSCTVVFGVMDNSPSVHLSGLIRCYQLVSRIDLKUYENFENIAMS 6270
Qy 110 ----- ELP ----- 112
Db 6271 IVIOKVLTANGLYPRIDSTGLCKHGFTTETIPSLRHPVSRGHALVFGANGFIGSIVFRLL 6330
Qy 113 ----- 112
Db 6331 QEMGMNVIPISRASIPSCDITNIKDVQNVFKSLGFKKFSVVVINCVGVETSAMKNTSLEQ 6390
Qy 113 ----- 112

```
Db 6391 EIVLSPKTEGYNILKCLEEFSIEVDKLVNFSSLSVWPLNGFDYASANCEVETLKQG 6450
Qy 113 -----DGNRQVRGV----- 122
Db 6451 SKYIKOFLTSLPPLGSRMYESTSTXHNLSQILMFOBELTNVLKTAFAESNMNGIVFV 6510
Qy 123 -----DGNRQVRGV----- 122
Db 6511 SSVNPADIATKSIAPHRDTENGSEVKAVESLKLPKSTCEVFIAEIKWETLGLISILNDAN 6570
Qy 123 -----QLGG----- 126
Db 6571 PNFSLGGDSLSALQVWVKVQKTDRIVDVNDLFDNPTLQEFKFKVKNLTTEKFAGTND 6630
Qy 127 -----A----- 126
Db 6631 KISYDAIPLNTSQTQNMRLQIDTTSKYNLIFKITISYETKFWBEFLKYLHSLIAYQPS 6690
Qy 127 -----ACSPT----- 131
Db 6691 YRTVFKSGNSPYQYICSLTESHPDKRCNLNNAISHPEPHLFEIGKSTPLRVRAEDCD 6750
Qy 132 -----WSCLITEDT----- 140
Db 6751 NSRIHIVFNQHHILTDGWSMTVLSDTVSLYAYRGETSFPSPKTKQTSQVAMGTSKSGD 6810
Qy 141 ----- 140
Db 6811 IKEALEYQNTYHTIIPYDSETGNTSPSYRISKIPSKIKWQKLVGLSKLYNTTMYNLAL 6870
Qy 141 ----- 140
Db 6871 SVFCDVRSFTGQADILLAYAISGRNADNSSELIGYFMNNAFKTSLPFELRLLEEILNIV 6930
Qy 141 -----G 141
Db 6931 LNSLEKSRGFATIPFYQMYEQNRKLNELSLFFNFROKLDYPTVSMFEGAKCEIEHLNLNA 6990
Qy 142 FDLGVTI 148
Db 6991 FDFSFTI 6997

RESULT 15
Q9L4X3
ID Q9L4X3 PRELIMINARY; PRT; 9477 AA.
AC Q9L4X3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ny61.
GN NY61.
OS Streptomyces noursei.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1971;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11455;
RX MEDLINE=20334850; PubMed=10873841;
RA Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA Valla S., Zotchev S.B.;
RT "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT deduction of the biosynthetic pathway."
RL Chem. Biol. 7:395-403 (2000).
DR EMBL; AF263912; AAF71766.1; -.
DR HSP; P81989; IQE5.
DR InterPro; IPR001227; AC transferase.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR004410; FAbD.
DR InterPro; IPR001899; Gram_pos_anchor.
```

```
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR001005; Myb_DNA_binding.
DR InterPro; IPR003880; Pantne_attach.
DR Pfam; PF00698; Acyl_transf; 6.
DR Pfam; PF00106; adh_short; 3.
DR Pfam; PF00109; ketoacyl-synt; 6.
DR Pfam; PF02801; ketoacyl-synt_C; 6.
DR Pfam; PF00550; pp-binding; 6.
DR TIGRFAMs; TIGR00128; fabD; 6.
DR PROSITE; PS50075; ACP_DOMAIN; 6.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 6.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE; PS00037; MYB_1; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 9477 AA; 988132 MW; 0D2BCA5D7B265483 CRC64;

Query Match 29 94; Score 231; DB 2; Length 9477;
Best Local Similarity 1.11; Pred.No. 57;
Matches 92; Conservative 27; Mismatches 29; Indels 7868; Gaps 29;

Qy 1 AAGGILH----- 7
Db 86 ASGFLHDPDFDADFFGCSPREAVAMPQQRVVLESWEAFERAGIDPTSVKGSRTGVF 145
Qy 8 ----- 7
Db 146 IGAMAQDYRVGPADGAEGFOLTGNTGTVLSGRISYTFCTGVPAVTVDTACSSSLVAVHLA 205
Qy 8 ----- 7
Db 206 TQALRAGECTLALAGGVTIMSGPOTFIEMGRQGLSADGRCSRCSFGDTADGTGWAEGVIL 265
Qy 8 ----- 7
Db 266 VLERLSDAVRNGHEILAVVGTAVNQDGSNGLTAPNGSQOQVQQALVNARLAAGDID 325
Qy 8 ----- 7
Db 326 VVEAHGTGTLGDPVEAQALLATYQGNRPADRPRLLLSGVSKNSLHSHTOAAAGVAGVIKVMV 385
Qy 8 ----- 7
Db 386 AMRHGTLPRTLHAEPTHVDSQGANVRLLTDTTDWPAWGAPRRRAVSVFSGISGTNAHTI 445
Qy 8 ----- 7
Db 446 IEQAPEPOPEDAATAQDDAAGSTPATAPVVGWVPVLLSGRTPDALRGQAAALRAALDTG 505
Qy 8 ----- 7
Db 506 RRPDLLDLAHSLATTPRAGFEHRAVLLATDHPALTDGLTALADADDPAAPAWITGTTRAE 565
Qy 8 ----- 7
Db 566 TRLAVLFTGQAGRLGAGRELAARPPAPATAALDAALDAFTPHLDRPLREVLWGTDAALLD 625
Qy 8 ----- 7
Db 626 RTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAHAHLAGVLSLADAATLVAARG 685
Qy 8 ----- 7
Db 686 RLNQALPDGAMIAVQASEADVAPLLAGHEDOVAIAAVNGSPSAVVLSGAEATVTALAEOL 745
Qy 8 ----- 7
Db 746 AADGRKTRRLRVSHAFSLMEPMLDAFRAVVEDLTLOPPLLPPVSVNLTKGPATVAQLTS 805
Qy 8 ----- 7
Db 806 ADYWDVHVHRAVRFADGIDMLARHDTTAFLELGPDGVLSAMAQDCLDAADADAVTLPALR 865
```

```
QY 8 ----- 7
Db 866 AGRPEHTLTATLHAGLHVHGTATLDMTCFAGTGARRTDLPTAFORRYWPKALQSGTAD 925
QY 8 -----LELL 11
Db 926 LRSVGLGAAHPLLSAAVSLADAGGTLTGRSLRQHPWLADHTVRGTTLLPGTAFLELA 985
QY 12 VAVG-----PDV----- 18
Db 986 VRAGDEVGCDRVEBELTLAAPLLPEQGGVQVQLWIGNPDVSGRRVTNVHARPDTGDDTPW 1045
QY 19 ----- 18
Db 1046 TAHATGVLTTADASROLPASSEQGTPLAGDPHPALDAAQWPPAGAEPLPLDGHYDRAD 1105
QY 19 ----- 18
Db 1106 GFGYGPVFOGLRAWRGGDVVYAEVELPEAGRSDAEAFGLHPALLDAAALHAAPTCLGE 1165
QY 19 ----- 18
Db 1166 RGRGLPFSMEGSLHAGGATTLRLVRLTPVADDALALTADGTGAPVLSVDSLVRSVAT 1225
QY 19 ----- 18
Db 1226 QOLDTAAAVARDALFRLDWTVPQATDPGPVALLGADPFGLLTHAGFADAPAYDPLAAL 1285
QY 19 ----- 18
Db 1286 AAADGPVTTVLSLGTGDDAADPARSAURCAEALAAVQTLDDHHERFAAARLVFVTR 1345
QY 19 ----- 18
Db 1346 GATVGRDVAAAANGLVRSQSPGCFALVDLPDGAAGAAALVAALVSGEPOLAVRGD 1405
QY 19 ----- 18
Db 1406 VLRVARLVRPLTEVGAGADGTGVDGSGVSFSGEGAVLVGTGGTGLGAVLARHLVAE 1465
QY 19 ----- 18
Db 1466 YGVRDLLLLRSRGERAVGAGELVAELAGVARGVRVACDVTDRAAVVELVGHGAVSAVH 1525
QY 19 ----- 18
Db 1526 AAGVLDDGMVGLTGERLSAVLRPKVDVAVHLEATRGDLDAFVFFSLAGVFGSQQA 1585
QY 19 ----- 18
Db 1586 NYAAANAFDALMTRRAEGLPGLSLANGPWEQSGGTGTLTDVDAERLARGVPPPLSVA 1645
QY 19 ----- 18
Db 1646 QGLALFDAAVAGTATCVPRDLPLVLRARGEVPLRLSLRVRARRAAVAGSATAGNLA 1705
QY 19 ----- 18
Db 1706 QRLRLDEDEGRMDVLVRGQVALVLGHATGQVDAGRAFRLDGLFDSLTAVELNRINT 1765
QY 19 ----- 18
Db 1766 VTGLRLPATLVFDYPTVRHLATYVLDELLGTDAEVATVQPAAVADDPIVIVGMACRYP 1825
QY 19 ----- 18
Db 1826 GCVSSPEDLWRVLTEGDAVSGFPTNRGMDVESLYHPDPDHGTSYTRSGGFLHEAGEFD 1885
QY 19 ----- 18
Db 1886 PGFGCMSPREALATDSQORLLLESWEAIERAGIDPVSLSRSRTGVGFAGVMYSYAML 1945
QY 19 -----FO----- 20
```

```
Db 1946 SPEFGFQSGSGSPSLASGRVAYTGLEGPVTVDTACSSSLVAMHMAQALRSCECGIA 2005
QY 21 ----- 20
Db 2006 LAGGVTMSTPAVDFVDFARQRLSPDGRCKAFADAADGVGMSEGVGLVLERQSDAVRNG 2065
QY 21 ----- 20
Db 2066 HEILAVRGSANQDGASNGLTAPNGPSOORVIRQALASGGITAGDQDVVVEAHGTGTTIG 2125
QY 21 ----- 20
Db 2126 DPTIAQALLATYGRORPERPRLLSGVSXNSLNGHTQAAAGVAGVVKMVLAMRHGVVPTLH 2185
QY 21 ----- 20
Db 2186 VDAPSSHVDWSEGAVELLSEQAAMPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAA 2245
QY 21 ----- 20
Db 2246 AGAARTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDABELVDVALSLATTSRFEQRA 2305
QY 21 ----- 20
Db 2306 AVVAQDRDQLIASLGLAADRDPAPVVEGEAAGRGRTAVLFTQGSORAAAGRELHEVOP 2365
QY 21 ----- 20
Db 2366 EFAAFDAVCAVDFPDLDRPLREVVFAPEDGDEAALLDETGTQTOPALFAVEVALFRLVES 2425
QY 21 ----- 20
Db 2426 WGVPRDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQATEDEI 2485
QY 21 -AHQEDT----- 26
Db 2486 AAHLDDTVAIAAVNGPQSVVISGDEEAETIAATFAERGRKTKRLRVSHAFHSPRMDGML 2545
QY 27 ----- 26
Db 2546 DAFRIVAEGLTVRAPRIPLVSDLTGRRADDAEVCTAEYVWRHVREAVRADCVRTLRDAG 2605
QY 27 ----- 26
Db 2606 ATTFELGSDGLLTAAEDTLGDDHDAELVPMLRAGRAEELAAATALARLOVRGVVDWA 2665
QY 27 ----- 26
Db 2666 AYLAGTGARRTDLPTVAFQHAYYWPQLPTPAALAAADPADQOLWAAVEMGDARELADIL 2725
QY 27 -----ERYVLTNL----- 34
Db 2726 GLGEODLTPLDSLPLALTSMRRGNQEKHLDDTLRYRVENWRLSKPTAPVLDDGTLLVASD 2785
QY 35 ----- 34
Db 2786 ATAADQALLDGLADALGSHGARVRRLLDDSCADRAVLAERLARTADVDAATQVLSVLP 2845
QY 35 ----- 34
Db 2846 LDERDADCPLTRGLALTVALVOALADTGAOGLMTATRGAVSTNPADPVTHPVQAAAW 2905
QY 35 ----- 34
Db 2906 GLORGVALEHPRLWGLVLDLPQVDFERAGORLAGILAVKADPDGEDOVALRATGVSGRRL 2965
QY 35 -----NIGAEELLR----- 42
Db 2966 VRHTEALPTAAEFTATGTVLTGGTGGLCAEVARMLARAGAOHLVTSRRGPDPAGAAE 3025
QY 43 -----FO----- 42
```

Db 3026 LRAELGCGPSVSVACDVADRDLAALVLTALPEELPLTGVVHTAGVGHYGLDTLSTAE 3085
Qy 43 ----- 42
Db 3086 FAGLTAAGLAGAHLADLADRELDFFVLFGSIAGVWGSNGOSAYGAANAYLDALALHRR 3145
Qy 43 -----DPS----- 45
Db 3146 ARGLAATSVAMGPWAEAGMAADDAVSETLRRQGLGLDPAPAMTELRRVAVRQDVTVTVA 3205
Qy 46 ----- 45
Db 3206 DVWQRYAPLFTSARPSALIAGLPEVRALAAADTEQDATGASEVVTRVRALAEPEQLRL 3265
Qy 46 ----- 45
Db 3266 LTDLVRTESATVLGHSSADAVEPGRAFRDVGFDSLTAVELRKRLGAATGLSLPSTWFDY 3325
Qy 46 -----LGAQFRV----- 52
Db 3326 PTPLELAQYLRAEILGAVLEAGVPATGADDEPIAIIQMCRFFPGGVSSPEQLMDLNAS 3385
Qy 53 ----- 52
Db 3386 GTDAISEFPVNRGWTGHLFDPPDRPGTTYSTQGGFLHEADFDPFTFGISPREALVMD 3445
Qy 53 ----- 52
Db 3446 POORLLLETTWESFERAGIRPETRLSTLTGTFGVSSYQBYGLGAGDGTGEGHMTGSSPSV 3505
Qy 53 -----HL----- 54
Db 3506 LSGRLSYVLEGPAVTDVTACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNFIA 3565
Qy 55 ----- 54
Db 3566 FSRQALAKDGRCKAFSDDADGMTLAEGVGVVLVERLSDAQNGHPVLAVLRGSAINQDG 3625
Qy 55 ----- 54
Db 3626 ASNGLTAPNGSQORVIRQALANARLAPGIDALEAHGTGTPLGDPIDAEQALFATYGRDR 3685
Qy 55 -----VQWV----- 58
Db 3686 DPESALLLGSVKSNIQTSQAAGIASVIKMMVALRHSLEPPTLHADAPSHVDWSAGTVR 3745
Qy 59 ILT----- 61
Db 3746 LLTQARAMPETGRPRRAAVSFGISGTNAHVILLEQAPVADTPAEBRPAVAPVPIAAGVVP 3805
Qy 62 ----- 61
Db 3806 WVTARSAALRGQAERLLAHAEVTGTALPAAGPLDIGLSLSARARFEHRAVVVPAGT 3865
Qy 62 ----- 61
Db 3866 DPLAALRAVATGPPSPVARGVADVEGRTVFVFPQGSQWGMGSQLLDES AVFAERIAE 3925
Qy 62 ----- 61
Db 3926 CAALAEFTDWSLDVLRGVGAPSLERVDVVQPASFAVMVSLAALWRSRGLVPDAVVGH 3985
Qy 62 ----- 61
Db 3986 SOGETAAAVVSGALSURDQARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGR 4045
Qy 62 -----EPE----- 64
Db 4046 VSVAAVNGPRSVVAGEPEALDALHRLTADDIRARRIAVDYASHHQVEDLHEELLEV 4105
Qy 65 ----- 64
Db 4106 AELAPRTSEVPFFSTVTGDWLDTARM DAGYWFNLRGRVRFADAVADLLAEYRAFVEYS 4165

Qy 65 ----- 64
Db 4166 SHPVLMSVAQEAIDEAGVPAAAGTLRRDOGGTDRFLLSAAAEVFRGVVDVWAGLFEFTG 4225
Qy 65 ----- 64
Db 4226 ASRIDLPTYAFOHEHLMWAVPPAPEAVAAADPDAAFAFTAVEDGDVSALTAAALGTDEDSVA 4285
Qy 65 ----- 64
Db 4286 AVLPAITSWRRARRDRSTVDARVRYVANKPLGGTLPHPSLTGCTWLLVTADGIDDDTDVAGA 4345
Qy 65 ----- 64
Db 4346 LETYGAEVRLVLDEECVDRVLRERLAGAEDVTGIVSVLAAAEERTDAVPGTSLVLGTAL 4405
Qy 65 ----- 64
Db 4406 TVALIQALGDAAIEDAPVWALTFRGAVSTGRADELTPAQVQTGIGWTAALHPQWGGTL 4465
Qy 65 ----- 64
Db 4466 DLPAALDARAAQRLAAVLVSGALGSDDDLAIIRPSGVFTTRIVRAEATAGRPAGTWTPRGTT 4525
Qy 65 ----- 64
Db 4526 LVTGSGGTLA PHLARWLAORGAELVLISRRGTAAPGAELVAELAESGTEATVAACDIT 4585
Qy 65 ----- 64
Db 4586 DRDAVALLADLKADGRTVRTVHTAATIELHTLDATTLADPDRVLHAKVTGAQVLAELL 4645
Qy 65 ----- 64
Db 4646 DDEELDDFVLISSTAGMWGGAHAAYVAGNAYLAALAEHRRANGLPALSLSWGIWADDLK 4705
Qy 65 ----- 64
Db 4706 LGRVDPQMIIRSGLEFMDPQLALSGLQALDDNNVLAVADVDMETHYHPVYTSGRPTPLF 4765
Qy 65 ----- 64
Db 4766 DEVPEVRLTAAAEQOSAGTVAEGEFAAALRALSDAEQOQTLLTETVTRTEAASVLGSSAED 4825
Qy 65 ----- 64
Db 4826 LTDQAFRDVGFDLSLTAVGLRNLASVTGLTLPSTWFDYPNPAAALAYLHGLSAGARSA 4885
Qy 65 ----- 64
Db 4886 AAGAAAVPTGAPDADDPTAIVGMSCRYPGVGSAAEDLWRIALDEVIDAISGFPADRGMDAE 4945
Qy 65 ----- 64
Db 4946 GLYDPPDRPGRYTSVQGGFLRDVAEFDGFFGISPREALSMDPQORLLLETAWAPEHA 5005
Qy 65 -----ITANLTS----- 75
Db 5006 GIDPVQGRSRTGTFTVGASYQDYASGVNPSGEGHMITGTLSVLSGRVSYLFGFEGPA 5065
Qy 76 ----- 75
Db 5066 VTLDTACSSSLVAMHLACQSLRNGESSLALAGGVSIMSTPMSPVGFSSRORALAEGRCKA 5125
Qy 76 ----- 75
Db 5126 YADGADGMTLAEGVGLVLLERLSARANGHOVLAVIRGSVAVNQDGASNGLTAPNGPSOOR 5185
Qy 76 -----SLLSVCG----- 82
Db 5186 VIRQALANSAPAGDIDVLEGHTGTALGDPIEAQALLATYQQDRAPERPLLLSGVKSNI 5245

QY	83	-----WSQTINP- 89
Db	5246	GHTOMASGVASVIKLVRALQEGVVPKSLHIDRPFSTHVDWSSGAIGLLTERTPWPETGRPR 5305
QY	90	----- 89
Db	5306	RAAVSSFGISGTVNHTILEQAPADEAPTADPPRDGLVPVLLSGRGEAALRAQAARLLAF 5365
QY	90	----- 89
Db	5366	VEERPEAHLTDLAHSLATSRAALERRAAVIAADRDTLRLGRALSDGRPDPLVQGTAGR 5425
QY	90	----- 89
Db	5426	GRTAFLFTGOGSORPGMGRHQRYPVFADALDEVLARLDGDRPLREVLFAPDSAEA 5485
QY	90	----- 89
Db	5486	ALLDRTGYAQPALFAVEVALFRLLTSWGLTPDVLGHSVGELEAAAHVAGVLSLDDACTLV 5545
QY	90	-----EDD----- 92
Db	5546	AARGRLMQALPEGAMVALEAAEDEVLPLEGLTDRVSAAVNGPRSVVVGVEEDVLL 5605
QY	93	----- 92
Db	5606	ADLFAADGRRTKRLRVSHAFSHPLMDAMLDDFAAVARGLTYHPTTIPFVSNVSGGLATAE 5665
QY	93	-----TDP----- 95
Db	5666	QVTPDPYVWGHVRAAVRFAOGIDWLATQGDVHTFLELPGDGLSAMARESLLTDPSTALL 5725
QY	96	----- 95
Db	5726	PTLGRDRPERPALVTAIAAAHAGHARVDMSGYPADHGARRTTLPTVAFQERYWPDTTAA 5785
QY	96	----- 95
Db	5786	TSAHTPGSALDABEFWAAVERDDVAALAASLDDATVMTVMVPAALTAWRRRRGGQTELDWS 5845
QY	96	----- 95
Db	5846	RYRVTKPRGGATAPAAALTCRWLVLPVPHDQRODDATAAADADETALGTTTVRLTVTT 5905
QY	96	----- 95
Db	5906	TDRAALAARITEAAGDQGPFGVLSLLPLATGDAGHPGAPAAALTLTTAVALGDAGIDA 5965
QY	96	----- 95
Db	5966	PLMNVTRGAVAVGRAEQVTAPQAAVWGLGRAVALELPARFGGTLDPATLDGOAARRLR 6025
QY	96	----- 95
Db	6026	AVLAATDGEDAVALRPSGVFLRLAHAPAGPDTARTAFDPAAGTVLITGGTGGIGGHVAR 6085
QY	96	----- 95
Db	6086	RLARDGATHLLTSRRGPAAPGADALRAELEELGARVTLAACDAADRDAALLAELPDD 6145
QY	96	----- 95
Db	6146	APLCAVHTAGVVEDHVVDALTPENFAAVLRAKTVAAHILHELTADLDLAFLFSSTAG 6205
QY	96	----- 95
Db	6206	VLGAAGOCGYAANAHLDALEHRRSHGLTALSVAWGPWAGSGMWADADELDRVRGGF 6265
QY	96	----- 95
Db	6266	EPLAPEPAVRALLRAIENDDTTVALADIWNERFORAFAAVRPLPFVADLPTETGRATPATA 6325
QY	96	-----GHAD----- 99

Db	6326	TGAATGLRQOALAEPEHERPAAVLDLLRTQVAAVLGHADPRTVEDDHAFRDLGFDLSLIL 6385
QY	100	-----LVLV----- 103
Db	6386	ELRNALNAATGLSLPATLVYDLPTPREMADFLLAELLGLTPTDTAATVASTASPKLSASF 6445
QY	104	-----ITRF----- 107
Db	6446	EOGGTPDDPIAVIGIGCRPGCVTTPEELWOLLDEGRDGISRFFDDRCWDUALCAGAS 6505
QY	108	----- 107
Db	6506	DTLEGGLTGVADFDARFFGISPREALAMPQORLLLETTWEALERAGIDPTTLRGSTTG 6565
QY	108	----- 107
Db	6566	VFVGTNGODYPTLLRRSASDVAGYVATGNTASVMSGRLSYALGLEGPVAVTIDTACSSLV 6625
QY	108	----- 107
Db	6626	ALHWAGRALRAGECDLVVAGGVSVMASPDSFVEFSTOGGLAPDGRCKAFSDAADGTAWSE 6685
QY	108	----- 107
Db	6686	GVGILLVLERLSAARNRHQVGLIRGTAVNQDGASNGLTAPNGLSQORVIAQALADARLR 6745
QY	108	-----DLEL----- 111
Db	6746	PADIDAIEAHGTGTLGDPPEARALITAYGRDRDAERPLLGLTVKSNIGHTQAAAGAAGV 6805
QY	112	----- 111
Db	6806	IKMLAMRHGTLPRTLHVGTPTSSHVDWGGTVALLDDARHPWPTGQPRRAGVSAPGVSGT 6865
QY	112	----- 111
Db	6866	NAHVVEQAPETEAAPAAEPAPATPTVVPVYVSGRSREALQALDRLTAHTAAHAPAR 6925
QY	112	-----PDGNRQ----- 117
Db	6926	SAADVGRSLATDRTLPHRAVLLAGPDGVREAAARAAAPRTPGRTAFLFSGOGAOLMHG 6985
QY	118	----- 117
Db	6986	DLYQRPVYADALDTVLAQFDTVLDPVPLRAALFAAPGTPEAALLDQGTGTPALFAVEVA 7045
QY	118	-----VRGVTQLGAGCS----- 129
Db	7046	LFRLAESHWRLTPDFVAGHSIGEIAAAHVAGVFSLEDACTLVAAARASLMQPRDGMVAL 7105
QY	130	----- 129
Db	7106	EATEDEVAPLLTDGVALAANGPRSVVWAGAEADVAVADRLAADGRRTRRLTTSVSHAFHS 7165
QY	130	----- 129
Db	7166	PLMDPMLTDFARVAEGLTYHEPRIPLVSTLLGAPAGAEURTDPYVVRHVRETVPFADGVR 7225
QY	130	-----PTW----- 129
Db	7226	ALHDAGAGTFVEIGPDGVLTALTQOTLDTVEAGAPAVVVPLORRDRBAGDLALLEGLATLH 7285
QY	130	----- 132
Db	7286	THGTGPSWPAFYFATGGHRTDLPTVAFQORYVWYVPELGAPVATAPQDPAARVYHETWAPLP 7345
QY	133	----- 132
Db	7346	APEAAAPAGRALVLPAGNRDRTAMMTAVADALGADTVTAEPAALABOLTAAGDTPWRVVV 7405
QY	133	----- 132

```
Db 7406 SLLAAASEGLPADGAWPAALLATLDEAGVHAPLWCVTRGAVAVAGEAPTAVGQAALWGLG 7465
Qy 133 ----- 132
Db 7466 RVAALDHPDRFCGLADLPADTDAAAGLLAAHAAAPGTEAEIATVATCGHARRLVRTPAA 7525
Qy 133 ----- 132
Db 7526 ADGATWLPGTGTVLVVGGTGTGCGRAARWLVRGARHLVLTAPDGTTTAADTEALTAE 7585
Qy 133 ----- 132
Db 7586 LAALGARITVVVDHPTAPDGFALLDGLPDDTPLTAVVYAPEADAAPGTAELSAAAPV 7645
Qy 133 ----- 134
Db 7646 TALGAALTGRPLDAFVLFGSIAGLVGRGAAEAAASGAYLDAFARACRDRGTPALAVWG 7705
Qy 135 ----- 134
Db 7706 AWADLVGPSLAAHLRMNGLPVMDADTALTALSRVADGSAAEAVADVWETFAFLHHEAR 7765
Qy 135 -----LITE----- 138
|:|
Db 7766 RTALFDALPEARALAEAAARDRAKTAAGDYGRWLAEQPAADHDATLLALVTEKAATVL 7825
Qy 139 ----- 138
Db 7826 GHADHDLLEPDLPRDLGFDLSLTAVDLRNQLTAETGLTLPATLVFDHPNPAALAAHLRAQ 7885
Qy 139 ----- 138
Db 7886 LLGEASPSAAPVAAPVALGADDDAIIVGMACRYPGGVTSPEDLWQLVGDEVDVGFPT 7945
Qy 139 DTGFDL----- 144
|:|
Db 7946 DRGHDLAALAGDGPGRSATAQGGFLYDATDFDGLFCISPREALVMDPQORILLETWEA 8005
Qy 145 ----- 144
Db 8006 LERAGIDPATLRGSGTTGVFVGGSGDYRPPPEAGQWQTAQSASLLSGRLAYTFGIQGT 8065
Qy 145 -----GVTI 148
|:|
Db 8066 VSVDTACSSSLVALHAAQALRAGECSIALAGGTV 8101
```

Search completed: March 20, 2003, 12:41:58
Job time : 273 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:34:52 ; Search time 37 Seconds
(without alignments)
533.002 Million cell updates/sec

Title: SEQID_1_15FUSED

Perfect score: 773

Sequence: 1 AAGGILHLELLVAVGPDPVQ.....SPTWSCLITDGTGDLGVTI 148

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.0

Searched: 908470 seqs, 13250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 101002:*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	773	100.0	148	23	Human Von Willebra
2	773	100.0	242	23	Human aggrecanase
3	773	100.0	933	23	Human ADAM-TS-like
4	773	100.0	1353	23	Human Von Willebra
5	773	100.0	1416	23	Human ADAMTS-M pol
6	773	100.0	1427	23	Human Von Willebra
7	770	99.6	203	22	Amino acid sequenc
8	770	99.6	1120	22	Amino acid sequenc
9	529	68.4	118	21	Human ORFX ORF2432
10	272	35.2	26926	22	Human titin (conne

11	264	34.2	15281	15	AAR44929
12	238	30.8	1784	21	AAB411379
13	237	30.7	997	22	AAB72283
14	237	30.7	1505	23	AAU72897
15	237	30.7	1686	22	AAB74944
16	237	30.7	1686	22	AAE00913
17	237	30.7	1686	22	AAE00934
18	236	30.5	11096	22	AAE10129
19	235	30.4	12199	21	AAE10180
20	231	29.9	1690	22	AAB86949
21	231	29.9	9477	22	AAE10144
22	229	29.6	2150	21	AAE10144
23	229	29.6	2165	22	AAE10144
24	226	29.2	5532	21	AAE10144
25	226	29.2	5532	22	AAE10144
26	224	29.0	6685	22	AAE10144
27	224	29.0	7107	22	AAE10144
28	224	29.0	9222	22	AAE10144
29	223	28.8	1054	22	AAE10144
30	216	27.9	4829	22	AAE10144
31	216	27.9	5588	20	AAE10144
32	216	27.9	5588	22	AAE10144
33	216	27.9	7068	22	AAE10144
34	216	27.9	7337	22	AAE10144
35	215	27.8	4260	22	AAE10144
36	215	27.8	4349	23	AAU79940
37	215	27.8	4349	23	ABB97540
38	215	27.8	4854	22	ABG06027
39	214	27.7	5985	23	AAU84799
40	214	27.7	6797	22	AAE10144
41	212	27.4	5035	13	AAE10144
42	212	27.4	5546	23	AAU85008
43	212	27.4	6619	22	ABG23329
44	211	27.3	4999	23	AAO22158
45	211	27.3	7718	22	ABG11811

ALIGNMENTS

RESULT 1

AAE24448
ID AAE24448 standard; Protein; 148 AA.

XX AC AAE24448;

XX DT 04-OCT-2002 (first entry)

XX DE Human Von Willebrand factor-cleaving protease fragment #1.

XX KW Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;
transgenic animal; immunisation; thromboembolic disease; preclampsia;
KW thrombotic thrombocytopenic purpura; TTP; Henoch-Schönlein purpura;
KW thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
transgenic; anticoagulant.

XX OS Homo sapiens.

XX PN WO200242441-A2.

XX PD 30-MAY-2002.

XX PF 20-NOV-2001; 2001WO-EPI3391.

XX PR 22-NOV-2000; 2000US-0721254.

XX PR 12-APR-2001; 2001US-0833328.

XX PA (BAXT) BAXTER AG.

XX PI Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
Scheiflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
PI Zimmermann K, Voelkel D;

T. niveum Cyclopo
Human ORFX ORF1143
Human ADAMTS-7 am
Human metalloprote
Human ADAM type me
Human 27875 ADAM-T
Human 27875 ADAM-T
Streptomyces nous
S. venezuelae pl
Human metalloprote
Streptomyces nous
Amino acid sequenc
Human secreted pro
S. avermitilis ave
Streptomyces averm
Novel human diagno
Drosophila melanog
Drosophila melanog
Human apoptosis-in
SpE a polypeptide
S. spinosa protein
Streptomyces nous
Novel human diagno
Human EST encoded
Human protocadheri
Novel human protei
Novel human diagno
HCV HepCia acrambl
Pimaricin biosynth
MH mutant porcine
Human melanocyte d
Novel human diagno
Ramoplanin bioeynt
Novel human diagno

```
DR WPI; 2002-479950/51.
XX N-PSDB; AAD39331.
XX Novel isolated or substantially purified Von Willebrand factor-cleaving
PT protease, useful for producing preparation for therapy of thrombosis
PT and thromboembolic disease such as thrombotic thrombocytopenic purpura
XX
XX Claim 1; Fig 3; 93pp; English.
XX
XX The invention relates to an isolated or substantially pure Von Willebrand
XX factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
XX purifying vWF which involves providing vWF-cp as a ligand, contacting a
XX solution comprising vWF with the polypeptide, ligand under conditions
XX where vWF is bound to the ligand and recovering from the ligand purified
XX vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
XX which involves immunising an animal with vWF-cp and isolating the anti-
XX vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
XX producing a preparation of prophylaxis and therapy of thrombosis and
XX thromboembolic disease such as thrombotic thrombocytopenic purpura (TTP),
XX Hemocho-Schonlein purpura, preeclampsia, neonatal thrombocytopenia or
XX haemolytic-uraemic syndrome. vWF-cp can also be used for processing
XX plasmatoc or recombinantly produced vWF. The invention is useful for
XX construction expression systems and generating transgenic animals which
XX express the polypeptide in vivo. The present sequence is human vWF-cp
XX fragment.
SQ Sequence 148 AA;
Query Match 100.0%; Score 773; DB 23; Length 148;
Best Local Similarity 100.0%; Pred. No. 3e-33;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGGILHLELLVAVGPDVFOAHQEDTERVLTNLNIGALLRDPGLGQAFRVHLVKMVL 60
Db 1 AAGGILHLELLVAVGPDVFOAHQEDTERVLTNLNIGALLRDPGLGQAFRVHLVKMVL 60
Qy 61 TEPEGAPNITANLTSSLSVCGWSQTIINPEDDTPGHADLVLYITRFDELDPGNRQVRG 120
Db 61 TEPEGAPNITANLTSSLSVCGWSQTIINPEDDTPGHADLVLYITRFDELDPGNRQVRG 120
Qy 121 VTQLGACSPWTSCLITEDTGFDLGVTI 148
Db 121 VTQLGACSPWTSCLITEDTGFDLGVTI 148
RESULT 2
AAU97641
XX ID AAU97641 standard; Protein; 242 AA.
XX AC AAU97641;
XX DT 27-AUG-2002 (first entry)
XX DE Human aggrecanase protein.
XX KW Aggrecanase; osteopathic; antiarthritic; antiinflammatory; human;
XX KW Cartilage; osteoarthritis; inflammatory disease; enzyme.
XX XX
XX OS Homo sapiens.
XX XX
XX PN WO200233093-A2.
XX XX
XX PD 25-APR-2002.
XX XX
XX PF 17-OCT-2001; 2001WO-US32458.
XX XX
XX PR 18-OCT-2000; 2000US-241469P.
XX XX
XX PA (GEMY ) GENETICS INST INC.
XX XX
XX PI Racie LA, Twine NC, Agostino MJ, Wolfman NM, Morris EA;
XX WPI; 2002-452389/48.
XX
```

```
DR N-PSDB; ABK52579.
XX Novel purified aggrecanase polypeptide useful for developing inhibitors
PT and antibodies to the aggrecanase polypeptide, which are useful for
PT treating aggrecanase-associated condition such as osteoarthritis
XX
XX Claim 1; Page 33-34; 41pp; English.
XX
XX This invention relates to the cDNA and protein sequences of a novel
XX human aggrecanase polypeptide. The protein of the invention may be
XX used to inhibit the proteolytic activity of aggrecanase, or to inhibit
XX the aggrecanase-mediated cleavage of aggrecan in cartilage. The
XX protein of the invention is useful for developing inhibitors of
XX aggrecanase protein. The cDNA sequence encoding the aggrecanase protein
XX of the invention is useful for designing probes for obtaining DNA
XX sequences encoding other aggrecanase molecules. The cDNA sequence is
XX also useful for detecting mRNA encoding aggrecanase in a given cell
XX population, and thus for detecting or diagnosing genetic disorders
XX involving the aggrecanase, or disorders involving cellular, organ or
XX tissue disorders in which aggrecanase is irregularly transcribed or
XX expressed. The DNA sequences may also be useful for preparing vectors
XX for gene therapy applications. An inhibitor of the protein is useful
XX in treating conditions characterised by degradation of articular
XX cartilage, by blocking the enzyme's proteolytic activity. An
XX aggrecanase protein inhibitor and a method for inhibition of its activity
XX are useful for treating various aggrecanase-associated conditions
XX including osteoarthritis and other inflammatory diseases. The
XX present sequence represents the human aggrecanase protein of the
XX invention.
SQ Sequence 242 AA;
Query Match 100.0%; Score 773; DB 23; Length 242;
Best Local Similarity 100.0%; Pred. No. 1e-32;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGGILHLELLVAVGPDVFOAHQEDTERVLTNLNIGALLRDPGLGQAFRVHLVKMVL 60
Db 42 AAGGILHLELLVAVGPDVFOAHQEDTERVLTNLNIGALLRDPGLGQAFRVHLVKMVL 101
Qy 61 TEPEGAPNITANLTSSLSVCGWSQTIINPEDDTPGHADLVLYITRFDELDPGNRQVRG 120
Db 102 TEPEGAPNITANLTSSLSVCGWSQTIINPEDDTPGHADLVLYITRFDELDPGNRQVRG 161
Qy 121 VTQLGACSPWTSCLITEDTGFDLGVTI 148
Db 162 VTQLGACSPWTSCLITEDTGFDLGVTI 189
RESULT 3
AAU79217
XX ID AAU79217 standard; Protein; 933 AA.
XX AC AAU79217;
XX DT 15-JUL-2002 (first entry)
XX DE Human ADAM-TS-like protein.
XX XX
XX KW Human; ADAM-TS-like protein; cardiovascular disorder; angina;
XX KW vascular system; congestive heart failure; myocardial infarction;
XX KW ischaemic heart disease; arrhythmia; hypertensive vascular disease;
XX KW secondary arterial hypertension; peripheral vascular disease; embolism;
XX KW chronic peripheral arterial occlusive disease; acute arterial thrombosis;
XX KW inflammatory vascular disorder; chronic obstructive pulmonary disease;
XX KW liver disorder.
XX XX
XX OS Homo sapiens.
XX XX
XX PN WO200226999-A2.
XX XX
XX PD 04-APR-2002.
XX XX
```

PF 26-SEP-2001; 2001WO-EP11124.
XX
PR 28-SEP-2000; 2000US-235881P.
PR 25-JUL-2001; 2001US-307333P.
XX
PA (FARB) BAYER AG.
XX
PI Xiao Y;
XX
DR WPI; 2002-383274/41.
DR N-PSDB; ABK493356.
XX
XX New purified human ADAM-TS-like protein, useful for identifying
PT modulators of protein activity for treating cardiovascular or liver
PT disorder or chronic obstructive pulmonary disease -
XX
XX Claim 25; Fig 2; 106pp; English.
XX
CC The invention relates to a human ADAM-TS-like protein and the
CC polynucleotide encoding it. The protein of the invention is useful for
CC treating cardiovascular disorders including diseases of the heart and
CC vascular system, such as congestive heart failure, myocardial infarction,
CC ischaemic heart diseases (e.g., stable angina, unstable angina), atrial
CC and ventricular arrhythmia, hypertensive vascular diseases (e.g., all
CC kinds of secondary arterial hypertension), and peripheral vascular
CC diseases (e.g., chronic peripheral arterial occlusive disease, acute
CC arterial thrombosis and embolism, inflammatory vascular disorders, etc),
CC liver disorders and chronic obstructive pulmonary disease. The sequences
CC are useful in diagnostic assays for detecting diseases and abnormalities
CC or susceptibility to diseases and abnormalities related to the presence
CC of mutations in the nucleic acid sequences which encode the protein. The
CC sequences are also useful for modulating ADAM-TS-like protein activity in
CC a disease condition. This sequence represents the human ADAM-TS-like
CC protein.
XX
SQ Sequence 933 AA;

Query Match 100.0%; Score 773; DB 23; Length 933;
Best Local Similarity 100.0%; Pred. No. 2.7e-31;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLRDPGLGAFRVHLVKMVL 60
Db 78 AAGGILHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLRDPGLGAFRVHLVKMVL 137

QY 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRG 120
Db 138 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRG 197

QY 121 VTOLGGACSPWTSCLITEDTGFDLGVTI 148
Db 198 VTOLGGACSPWTSCLITEDTGFDLGVTI 225

RESULT 4
AAE24449
ID AAE24449 standard; Protein; 1353 AA.
XX
AC AAE24449;
XX
XX 04-OCT-2002 (first entry)
XX
DE Human Von Willebrand factor-cleaving protease fragment #2.
XX
KW Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;
KW transgenic animal; immunisation; thromboembolic disease; preclampsia;
KW thrombotic thrombocytopenic purpura; TTP; Henoch-Schonlein purpura;
KW thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
KW transgenic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO200242441-A2.

XX
PD 30-MAY-2002.
XX
PF 20-NOV-2001; 2001WO-EPI13391.
XX
XX 22-NOV-2000; 2000US-0721254.
PR 12-APR-2001; 2001US-0833328.
XX
PA (BAXT) BAXTER AG.
XX
XX Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
PI Scheifflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
PI Zimmermann K, Voelkel D;
XX
XX WPI; 2002-479950/51.
XX
PT Novel isolated or substantially purified Von Willebrand factor-cleaving
PT protease, useful for producing preparation for therapy of thrombosis
PT and thromboembolic disease such as thrombotic thrombocytopenic purpura -
XX
XX Claim 1; Page 64-68; 93pp; English.
XX
CC The invention relates to an isolated or substantially pure Von Willebrand
CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
CC solution comprising vWF with the polypeptide ligand under conditions
CC where vWF is bound to the ligand and recovering from the ligand purified
CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
CC which involves immunising an animal with vWF-cp and isolating the anti-
CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
CC producing a preparation of prophylaxis and therapy of thrombosis and
CC thromboembolic disease such as thrombotic thrombocytopenic purpura (TTP),
CC Henoch-Schonlein purpura, preclampsia, neonatal thrombocytopenia or
CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
CC plasmatic or recombinantly produced vWF. The invention is useful for
CC construction expression systems and generating transgenic animals which
CC express the polypeptide in vivo. The present sequence is human vWF-cp
CC fragment.
XX
SQ Sequence 1353 AA;

Query Match 100.0%; Score 773; DB 23; Length 1353;
Best Local Similarity 100.0%; Pred. No. 6.7e-31;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLRDPGLGAFRVHLVKMVL 60
Db 1 AAGGILHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLRDPGLGAFRVHLVKMVL 60

QY 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRG 120
Db 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRG 120

QY 121 VTOLGGACSPWTSCLITEDTGFDLGVTI 148
Db 121 VTOLGGACSPWTSCLITEDTGFDLGVTI 148

RESULT 5
ABB04153
ID ABB04153 standard; protein; 1416 AA.
XX
AC ABB04153;
XX
XX 26-MAR-2002 (first entry)
DT
XX Human ADAMTS-M polypeptide.
DE
XX Osteoarthritis; rheumatoid arthritis; inflammatory bowel disease;
KW Crohn's disease; asthma; Alzheimer's disease; organ transplant rejection;
KW cachexia; allergy; cancer; leukaemia; lymphoma; osteoporosis;
KW atherosclerosis; congestive heart failure; myocardial infarction; stroke;
KW neurodegenerative disease; autoimmune disorder; Huntington's;

KW Parkinson's; migraine; pain; depression; multiple sclerosis; burn;
 KW infertility; diabetic shock; gene therapy; ADAMTS-M;
 KW A Disintegrin And Metalloprotease; thrombospondin domain.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Peptide 1..97
 FT /label= Prodomain
 FT /note= "The mature form of the ADAMTS-M protein is
 FT processed by furin cleavage of the prodomain"
 FT 94..97
 FT /label= Furin_cleavage_site
 FT Protein 98..1416
 FT /label= Mature_ADAMTS-M_protein
 FT /note= "The mature form of the ADAMTS-M protein is
 FT processed by furin cleavage of the prodomain"
 FT 98..311
 FT /label= Metalloprotease_domain
 FT Domain 247..272
 FT /label= Zinc-binding_motif
 FT Domain 324..394
 FT /label= Disintegrin_domain
 FT Domain 410..473
 FT /label= Thrombospondin_submotif
 FT Domain 419..424
 FT /label= Heparin-binding_domain
 FT Domain 1099..1156
 FT /label= Thrombospondin_submotif
 XX
 PN EP1152055-A1.
 XX
 PD 07-NOV-2001.
 XX
 XX 24-APR-2001; 2001EP-0303708.
 XX
 PF 27-APR-2000; 2000US-200040P.
 XX
 PR (PFIZ) PFIZER PROD INC.
 XX
 PA Buckbinder L, Mitchell PG, Wachtmann TS, Walsh RT;
 PI
 PI WPI; 2002-084275/12.
 DR N-PSDB; ABA02549.
 DR
 XX
 PT New polynucleotide, useful in gene therapy, particularly for treating
 PT or preventing e.g. arthritis, Crohn's disease, Alzheimer's disease and
 PT organ transplant toxicity and rejection, comprises ADAMTS
 PT polynucleotide and encoded polypeptide -
 XX
 PS Claim 4; Fig 2; 31pp; English.
 XX
 CC The present sequence represents a ADAMTS protein, designated ADAMTS-M,
 CC that exhibits the characteristics of the ADAM (A Disintegrin And
 CC Metalloprotease) family of metalloproteases, and contains a
 CC thrombospondin domain (TS). The protein is encoded by the cDNA given in
 CC ABA02549. The specification describes a newly isolated polynucleotide,
 CC comprising a nucleotide sequence encoding an ADAMTS-M polypeptide as
 CC given in the specification, or a metalloproteinase, disintegrin domain,
 CC prodomain or its thrombospondin submotif. The polynucleotide, polypeptide
 CC and agent are useful for manufacturing a medicament for treating a
 CC subject in need of altering activity or expression of ADAMTS-M. The
 CC polynucleotide, ADAMTS-M polypeptide and agent are useful for
 CC manufacturing a medicament for treating arthritis (osteoarthritis and
 CC rheumatoid arthritis), inflammatory bowel disease, Crohn's disease,
 CC asthma, Alzheimer's disease, organ transplant cancer including rejection,
 CC cachexia, allergy, cancer (e.g. solid tumour malignancies including
 CC breast, lung, prostate, brain or haematopoietic malignancies including
 CC leukaemia and lymphoma), osteoporosis, atherosclerosis, aortic aneurysm,
 CC congestive heart failure, myocardial infarction, stroke, head trauma,
 CC spinal cord injury, neurodegenerative disease, autoimmune disorders,
 CC Huntington's disease, Parkinson's disease, migraine, pain, depression,
 CC multiple sclerosis, abnormal wound healing, burns, infertility or

CC diabetic shock. The polynucleotide and polypeptide are also useful for
 CC diagnosing the diseases above. The polynucleotide is particularly useful
 CC in gene therapy for treating the diseases cited above.
 XX
 SQ Sequence 1416 AA;
 Query Match 100.0%; Score 773; DB 23; Length 1416;
 Best Local Similarity 100.0%; Pred. No. 7.5e-31;
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AAGGILHLELLVAVGPDVFOAQHOEDTERRYVLTNLTGAEALLRDPISLGAQFRVHLVKMVL 60
 DB 98 AAGGILHLELLVAVGPDVFOAQHOEDTERRYVLTNLTGAEALLRDPISLGAQFRVHLVKMVL 157
 OY 61 TEPEGAPNITANTSSLLSVCWSQINPEDDDTDPGHADLVLYITRFDLELPDGNRQVRG 120
 DB 158 TEPEGAPNITANTSSLLSVCWSQINPEDDDTDPGHADLVLYITRFDLELPDGNRQVRG 217
 OY 121 VTOLGGACSPWTSCLITEDTGFDLGVTI 148
 DB 218 VTOLGGACSPWTSCLITEDTGFDLGVTI 245
 RESULT 6
 ID AAE24450
 XX AAE24450 standard; Protein; 1427 AA.
 AC AAE24450;
 XX
 DT 04-OCT-2002 (first entry)
 XX Human Von Willebrand factor-cleaving protease (vWF-cp).
 DE
 XX Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;
 KW transgenic animal; immunisation; thromboembolic disease; preeclampsia;
 KW thrombotic thrombocytopenic purpura; TTP; Henoch-Schönlein purpura;
 KW thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
 KW transgenic; anticoagulant; chromosome 9.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Peptide 1..51
 FT /label= Signal_peptide
 FT Protein 52..1427
 FT /note= "Mature human vWF-cp protein"
 FT Cleavage-site 69..75
 FT /note= "Furin cleavage site"
 FT Region 224..228
 FT /note= "Catalytic side"
 FT Region 249
 FT /note= "Met turn"
 FT Domain 301..377
 FT /note= "Disintegrin like motif"
 FT Domain 387..439
 FT /note= "Thrombospondin type I motif"
 FT Region 441..553
 FT /note= "Cysteine rich region"
 FT Region 554..687
 FT /note= "Spacer"
 FT Domain 688..743
 FT /note= "Thrombospondin type I motif"
 FT Domain 744..805
 FT /note= "Thrombospondin type I motif"
 FT Domain 897..952
 FT /note= "Thrombospondin type I motif"
 FT Domain 953..1013
 FT /note= "Thrombospondin type I motif"
 FT Domain 1016..1073
 FT /note= "Thrombospondin type I motif"
 FT Domain 1075..1131
 FT /note= "Thrombospondin type I motif"
 XX

PN W0200242441-A2.
XX 30-MAY-2002.
XX 20-NOV-2001; 2001WO-EP13391.
XX 22-NOV-2000; 2000US-0721254.
PR 12-APR-2001; 2001US-083328.
XX (BAXT) BAXTER AG.
XX Laemle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
PI Scheiflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
PI Zimmermann K, Voelkel D;
XX WPI: 2002-479950/51.
DR N-PSDB; AAD39332.
XX Novel isolated or substantially purified Von Willebrand factor-cleaving
PT protease, useful for producing preparation for therapy of thrombosis
PT and thromboembolic disease such as thrombotic thrombocytic purpura -
XX Claim 1: Fig 5; 93pp; English.
XX The invention relates to an isolated or substantially pure Von Willebrand
CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
CC solution comprising vWF with the polypeptide ligand under conditions
CC where vWF is bound to the ligand and recovering from the ligand purified
CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
CC which involves immunising an animal with vWF-cp and isolating the anti-
CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
CC producing a preparation of prophylaxis and therapy of thrombosis and
CC thromboembolic disease such as thrombotic thrombocytic purpura (TTP),
CC Henoch-Schönlein purpura, preeclampsia, neonatal thrombocytopenia or
CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
CC plasmatic or recombinantly produced vWF. The invention is useful for
CC constriction expression systems and generating transgenic animals which
CC express the polypeptide in vivo. The present sequence is human vWF-cp
CC protein. vWF-cp gene is located on chromosome 9.
XX Sequence 1427 AA;
SQ Query Match 100.0%; Score 773; DB 23; Length 1427;
Best Local Similarity 100.0%; Pred. No. 7.7e-31;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AAGGILHLELLVAVGPDVFOAQHEDTERVYLTNLNIGALLRDPSPSLCAQFRVHLVKWVIL 60
DB 75 AAGGILHLELLVAVGPDVFOAQHEDTERVYLTNLNIGALLRDPSPSLCAQFRVHLVKWVIL 134
OY 61 TEPEGAPNITANLTSSLSVCGWSQINPEDDTPDGHADLVLYITRFDLELDPGNRQVRG 120
DB 135 TEPEGAPNITANLTSSLSVCGWSQINPEDDTPDGHADLVLYITRFDLELDPGNRQVRG 194
OY 121 VTOLGGACSPWNSCLITDGTGFDLGVTI 148
DB 195 VTOLGGACSPWNSCLITDGTGFDLGVTI 222
RESULT 7
AAG63826
ID AAG63826 standard; Protein; 203 AA.
XX AAG63826;
AC
XX
DT 29-OCT-2001 (first entry)
XX Amino acid sequence of a human zDint5 polypeptide.
XX Human; zDint5; anti-angiogenic; intestinal polypeptide; wound healing;
KW extracellular matrix interaction; tumour suppression; gamete maturation;
KW immunologic recognition; gastrointestinal irradiation; chemotherapy;

KW proteolysis; apoptosis; angiogenesis; infection; cell adhesion;
KW cell fusion; cell signalling; tumour; Crohn's disease; melanoma; trauma;
KW inflammatory bowel disease; food poisoning; degenerative disease;
KW inflammation; fertility; gamete maturation; epithelial disorder.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH Modified-site 69
FT /note= "potential N-linked glycosylation site"
FT Modified-site 73
FT /note= "potential N-linked glycosylation site"
FT Region 151..161
FT /note= "zinc-binding motif"
XX W0200159112-A1.
XX 16-AUG-2001.
PD
XX 09-FEB-2001; 2001WO-US04198.
PF
XX 10-FEB-2000; 2000US-0501806.
PR
XX (ZYMO) ZYMOGENETICS INC.
PA Holloway JL, Sheppard PO;
XX
PI WPI: 2001-522477/57.
DR N-PSDB; AAH74759.
XX New anti-angiogenic intestinal polypeptides, zDint5 polypeptides, which
PT are members of disintegrin proteases, for modulating extracellular
PT matrix interaction, tumour suppression and wound healing -
XX Claim 1: Page 73-74; 92pp; English.
XX The present sequence represents a human zDint5 polypeptide. The zDint5
CC polypeptide is an anti-angiogenic intestinal polypeptide. zDint5 is
CC used for modulating extracellular matrix interactions. zDint5
CC polypeptide is useful as a tool for identifying new family members
CC of polypeptides. zDint5 polynucleotides are useful as probes or primers
CC to clone 5' non-coding regions of zDint5 gene. zDint5 polypeptides are
CC used for tumour suppression, gamete maturation, immunologic recognition,
CC and growth and differentiation either working in isolation or in
CC conjunction with other molecules in colon, small intestine, fetal lung,
CC testis and B-cells. zDint5 polypeptides are also useful for promoting
CC wound healing, in the treatment of disorders associated with recovery
CC after gastrointestinal irradiation, chemotherapy or antibody use.
CC as anti-infectives, and extracellular matrix repair and remodeling. The
CC polypeptides are also useful for modulating proteolysis, apoptosis,
CC angiogenesis, infection, cell adhesion, cell fusion and signalling.
CC The polypeptides are also useful for treating tumour formation, Crohn's
CC disease, inflammatory bowel disease, food poisoning, melanoma,
CC degenerative diseases, disorders related to immunity, inflammation,
CC fertility, gamete maturation, immunology, trauma and epithelial
CC disorders.
XX
SQ Sequence 203 AA;
Query Match 99.6%; Score 770; DB 22; Length 203;
Best Local Similarity 99.3%; Pred. No. 9.4e-33;
Matches 147; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 AAGGILHLELLVAVGPDVFOAQHEDTERVYLTNLNIGALLRDPSPSLCAQFRVHLVKWVIL 60
DB 2 AAGGILHLELLVAVGPDVFOAQHEDTERVYLTNLNIGALLRDPSPSLCAQFRVHLVKWVIL 61
OY 61 TEPEGAPNITANLTSSLSVCGWSQINPEDDTPDGHADLVLYITRFDLELDPGNRQVRG 120
DB 62 TEPEGAPNITANLTSSLSVCGWSQINPEDDTPDGHADLVLYITRFDLELDPGNRQVRG 121
OY 121 VTOLGGACSPWNSCLITDGTGFDLGVTI 148
|||||

Db	122 VTQGGACSPWWSCLITDGTGFDLGVTI 149	CC	after gastrointestinal irradiation, chemotherapy or antibody use,
		CC	as anti-infectives, and extracellular matrix repair and remodeling. The
		CC	polypeptides are also useful for modulating proteolysis, apoptosis,
		CC	angiogenesis, infection, cell adhesion, cell fusion and signalling.
		CC	The polypeptides are also useful for treating tumour formation, Crohn's
		CC	disease, inflammatory bowel disease, food poisoning, melanoma,
		CC	degenerative diseases, disorders related to immunity, inflammation,
		CC	fertility, gamete maturation, immunology, trauma and epithelial
		CC	disorders.
XX		XX	
SO	Sequence 1120 AA;	SO	
			Query Match 99.6%; Score 770; DB 22; Length 1120;
			Best Local Similarity 99.3%; Pred. No. 6.1e-31;
			Matches 147; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY	1 AAGGIIHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLDDPSLGAFRVHLVKWVIL 60	OY	
Db	105 AAGGIIHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLDDPSLGAFRVHLVKWVIL 164	Db	
OY	61 TEPEGAPNITANLTSSLLSVCGWSQINPEDDTPDGHADLVLYITRFDLELPDGNRQVRG 120	OY	
Db	165 TEPEGAPNITANLTSSLLSVCGWSQINPEDDTPDGHADLVLYITRFDLELPDGNRQVRG 224	Db	
OY	121 VTQGGACSPWWSCLITDGTGFDLGVTI 148	OY	
Db	225 VTQGGACSPWWSCLITDGTGFDLGVTI 252	Db	
			RESULT 9
			AAB42668
ID	AAB42668 standard; Protein; 118 AA.	ID	
XX		XX	
AC	AAB42668;	AC	
DT		DT	
XX	08-FEB-2001 (first entry)	XX	
XX		XX	
DE	Human ORFX ORF2432 polypeptide sequence SEQ ID NO:4864.	DE	
XX		XX	
KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;	KW	
KW	vulvular; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;	KW	
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;	KW	
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;	KW	
KW	hypotensive; dermatological; immunosuppressive; antitumour;	KW	
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;	KW	
KW	antianemic; gene therapy; cancer; proliferative disorder; hypertension;	KW	
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;	KW	
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;	KW	
KW	cholesterol ester storage; systemic lupus erythematosus; infection;	KW	
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;	KW	
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;	KW	
KW	bone damage; cartilage damage; antinflammatory disease; coagulation;	KW	
KW	thrombosis; contraceptive.	KW	
XX		XX	
OS	Homo sapiens.	OS	
XX		XX	
PN	WO200058473-A2.	PN	
XX		XX	
PD	05-OCT-2000.	PD	
XX		XX	
PF	31-MAR-2000; 2000WO-US08621.	PF	
XX		XX	
PR	31-MAR-1999; 99US-0127607.	PR	
PR	02-APR-1999; 99US-0127636.	PR	
PR	05-APR-1999; 99US-0127728.	PR	
XX	30-MAR-2000; 2000US-0540763.	XX	
PA	(CURA-) CURAGEN CORP.	PA	
XX		XX	
PI	Shankers RA, Leach M;	PI	
XX		XX	
DR	WPI; 2000-602362/57.	DR	
DR	N-PSDB; AAC76877.	DR	

Db	122 VTQGGACSPWWSCLITDGTGFDLGVTI 149	CC	after gastrointestinal irradiation, chemotherapy or antibody use,
		CC	as anti-infectives, and extracellular matrix repair and remodeling. The
		CC	polypeptides are also useful for modulating proteolysis, apoptosis,
		CC	angiogenesis, infection, cell adhesion, cell fusion and signalling.
		CC	The polypeptides are also useful for treating tumour formation, Crohn's
		CC	disease, inflammatory bowel disease, food poisoning, melanoma,
		CC	degenerative diseases, disorders related to immunity, inflammation,
		CC	fertility, gamete maturation, immunology, trauma and epithelial
		CC	disorders.
XX		XX	
SO	Sequence 1120 AA;	SO	
			Query Match 99.6%; Score 770; DB 22; Length 1120;
			Best Local Similarity 99.3%; Pred. No. 6.1e-31;
			Matches 147; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY	1 AAGGIIHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLDDPSLGAFRVHLVKWVIL 60	OY	
Db	105 AAGGIIHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLDDPSLGAFRVHLVKWVIL 164	Db	
OY	61 TEPEGAPNITANLTSSLLSVCGWSQINPEDDTPDGHADLVLYITRFDLELPDGNRQVRG 120	OY	
Db	165 TEPEGAPNITANLTSSLLSVCGWSQINPEDDTPDGHADLVLYITRFDLELPDGNRQVRG 224	Db	
OY	121 VTQGGACSPWWSCLITDGTGFDLGVTI 148	OY	
Db	225 VTQGGACSPWWSCLITDGTGFDLGVTI 252	Db	
			RESULT 9
			AAB42668
ID	AAB42668 standard; Protein; 118 AA.	ID	
XX		XX	
AC	AAB42668;	AC	
DT		DT	
XX	08-FEB-2001 (first entry)	XX	
XX		XX	
DE	Human ORFX ORF2432 polypeptide sequence SEQ ID NO:4864.	DE	
XX		XX	
KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;	KW	
KW	vulvular; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;	KW	
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;	KW	
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;	KW	
KW	hypotensive; dermatological; immunosuppressive; antitumour;	KW	
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;	KW	
KW	antianemic; gene therapy; cancer; proliferative disorder; hypertension;	KW	
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;	KW	
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;	KW	
KW	cholesterol ester storage; systemic lupus erythematosus; infection;	KW	
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;	KW	
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;	KW	
KW	bone damage; cartilage damage; antinflammatory disease; coagulation;	KW	
KW	thrombosis; contraceptive.	KW	
XX		XX	
OS	Homo sapiens.	OS	
XX		XX	
PN	WO200058473-A2.	PN	
XX		XX	
PD	05-OCT-2000.	PD	
XX		XX	
PF	31-MAR-2000; 2000WO-US08621.	PF	
XX		XX	
PR	31-MAR-1999; 99US-0127607.	PR	
PR	02-APR-1999; 99US-0127636.	PR	
PR	05-APR-1999; 99US-0127728.	PR	
XX	30-MAR-2000; 2000US-0540763.	XX	
PA	(CURA-) CURAGEN CORP.	PA	
XX		XX	
PI	Shankers RA, Leach M;	PI	
XX		XX	
DR	WPI; 2000-602362/57.	DR	
DR	N-PSDB; AAC76877.	DR	

Db	122 VTQGGACSPWWSCLITDGTGFDLGVTI 149	CC	after gastrointestinal irradiation, chemotherapy or antibody use,
		CC	as anti-infectives, and extracellular matrix repair and remodeling. The
		CC	polypeptides are also useful for modulating proteolysis, apoptosis,
		CC	angiogenesis, infection, cell adhesion, cell fusion and signalling.
		CC	The polypeptides are also useful for treating tumour formation, Crohn's
		CC	disease, inflammatory bowel disease, food poisoning, melanoma,
		CC	degenerative diseases, disorders related to immunity, inflammation,
		CC	fertility, gamete maturation, immunology, trauma and epithelial
		CC	disorders.
XX		XX	
SO	Sequence 1120 AA;	SO	
			Query Match 99.6%; Score 770; DB 22; Length 1120;
			Best Local Similarity 99.3%; Pred. No. 6.1e-31;
			Matches 147; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY	1 AAGGIIHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLDDPSLGAFRVHLVKWVIL 60	OY	
Db	105 AAGGIIHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLDDPSLGAFRVHLVKWVIL 164	Db	
OY	61 TEPEGAPNITANLTSSLLSVCGWSQINPEDDTPDGHADLVLYITRFDLELPDGNRQVRG 120	OY	
Db	165 TEPEGAPNITANLTSSLLSVCGWSQINPEDDTPDGHADLVLYITRFDLELPDGNRQVRG 224	Db	
OY	121 VTQGGACSPWWSCLITDGTGFDLGVTI 148	OY	
Db	225 VTQGGACSPWWSCLITDGTGFDLGVTI 252	Db	
			RESULT 9
			AAB42668
ID	AAB42668 standard; Protein; 118 AA.	ID	
XX		XX	
AC	AAB42668;	AC	
DT		DT	
XX	08-FEB-2001 (first entry)	XX	
XX		XX	
DE	Human ORFX ORF2432 polypeptide sequence SEQ ID NO:4864.	DE	
XX		XX	
KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;	KW	
KW	vulvular; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;	KW	
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;	KW	
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;	KW	
KW	hypotensive; dermatological; immunosuppressive; antitumour;	KW	
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;	KW	
KW	antianemic; gene therapy; cancer; proliferative disorder; hypertension;	KW	
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;	KW	
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;	KW	
KW	cholesterol ester storage; systemic lupus erythematosus; infection;	KW	
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;	KW	
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;	KW	
KW	bone damage; cartilage damage; antinflammatory disease; coagulation;	KW	
KW	thrombosis; contraceptive.	KW	
XX		XX	
OS	Homo sapiens.	OS	
XX		XX	
PN	WO200058473-A2.	PN	
XX		XX	
PD	05-OCT-2000.	PD	
XX		XX	
PF	31-MAR-2000; 2000WO-US08621.	PF	
XX		XX	
PR	31-MAR-1999; 99US-0127607.	PR	
PR	02-APR-1999; 99US-0127636.	PR	
PR	05-APR-1999; 99US-0127728.	PR	
XX	30-MAR-2000; 2000US-0540763.	XX	
PA	(CURA-) CURAGEN CORP.	PA	
XX		XX	
PI	Shankers RA, Leach M;	PI	
XX		XX	
DR	WPI; 2000-602362/57.	DR	
DR	N-PSDB; AAC76877.	DR	

Db	122 VTQGGACSPWWSCLITDGTGFDLGVTI 149	CC	after gastrointestinal irradiation, chemotherapy or antibody use,
		CC	as anti-infectives, and extracellular matrix repair and remodeling. The
		CC	polypeptides are also useful for modulating proteolysis, apoptosis,
		CC	angiogenesis, infection, cell adhesion, cell fusion and signalling.
		CC	The polypeptides are also useful for treating tumour formation, Crohn's
		CC	disease, inflammatory bowel disease, food poisoning, melanoma,
		CC	degenerative diseases, disorders related to immunity, inflammation,
		CC	fertility, gamete maturation, immunology, trauma and epithelial
		CC	disorders.
XX		XX	
SO	Sequence 1120 AA;	SO	
			Query Match 99.6%; Score 770; DB 22; Length 1120;
			Best Local Similarity 99.3%; Pred. No. 6.1e-31;
			Matches 147; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY	1 AAGGIIHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLDDPSLGAFRVHLVKWVIL 60	OY	
Db	105 AAGGIIHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLDDPSLGAFRVHLVKWVIL 164	Db	
OY	61 TEPEGAPNITANLTSSLLSVCGWSQINPEDDTPDGHADLVLYITRFDLELPDGNRQVRG 120	OY	
Db	165 TEPEGAPNITANLTSSLLSVCGWSQINPEDDTPDGHADLVLYITRFDLELPDGNRQVRG 224	Db	
OY	121 VTQGGACSPWWSCLITDGTGFDLGVTI 148	OY	
Db	225 VTQGGACSPWWSCLITDGTGFDLGVTI 252	Db	
			RESULT 9
			AAB42668
ID	AAB42668 standard; Protein; 118 AA.	ID	
XX		XX	
AC	AAB42668;	AC	
DT		DT	
XX	08-FEB-2001 (first entry)	XX	
XX		XX	
DE	Human ORFX ORF2432 polypeptide sequence SEQ ID NO:4864.	DE	
XX		XX	
KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;	KW	
KW	vulvular; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;	KW	
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;	KW	
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;	KW	
KW	hypotensive; dermatological; immunosuppressive; antitumour;	KW	
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;	KW	
KW	antianemic; gene therapy; cancer; proliferative disorder; hypertension;	KW	
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;	KW	
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;	KW	
KW	cholesterol ester storage; systemic lupus erythematosus; infection;	KW	
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;	KW	
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;	KW	
KW	bone damage; cartilage damage; antinflammatory disease; coagulation;	KW	
KW	thrombosis; contraceptive.	KW	
XX		XX	
OS	Homo sapiens.	OS	
XX		XX	
PN	WO200058473-A2.	PN	
XX		XX	
PD	05-OCT-2000.	PD	
XX		XX	
PF	31-MAR-2000; 2000WO-US08621.	PF	
XX		XX	
PR	31-MAR-1999; 99US-0127607.	PR	
PR	02-APR-1999; 99US-0127636.	PR	
PR	05-APR-1999; 99US-0127728.	PR	
XX	30-MAR-2000; 2000US-0540763.	XX	
PA	(CURA-) CURAGEN CORP.	PA	
XX		XX	
PI	Shankers RA, Leach M;	PI	
XX		XX	
DR	WPI; 2000-602362/57.	DR	
DR	N-PSDB; AAC76877.	DR	

Db	122 VTQGGACSPWWSCLITDGTGFDLGVTI 149	CC	after gastrointestinal irradiation, chemotherapy or antibody use,
		CC	as anti-infectives, and extracellular matrix repair and remodeling. The
		CC	polypeptides are also useful for modulating proteolysis, apoptosis,
		CC	angiogenesis, infection, cell adhesion, cell fusion and signalling.
		CC	The polypeptides are also useful for treating tumour formation, Crohn's
		CC	disease, inflammatory bowel disease, food poisoning, melanoma,
		CC	degenerative diseases, disorders related to immunity, inflammation,
		CC	fertility, gamete maturation, immunology, trauma and epithelial
		CC	disorders.
XX		XX	
SO	Sequence 1120 AA;	SO	
			Query Match 99.6%; Score 770; DB 22; Length 1120;
			Best Local Similarity 99.3%; Pred. No. 6.1e-31;
			Matches 147; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY	1 AAGGIIHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLDDPSLGAFRVHLVKWVIL 60	OY	
Db	105 AAGGIIHLELLVAVGPDVFOAQHEDTERVLTNLNIGALLDDPSLGAFRVHLVKWVIL 164	Db	
OY	61 TEPEGAPNITANLTSSLLSVCGWSQINPEDDTPDGHADLVLYITRFDLELPDGNRQVRG 120	OY	
Db	165 TEPEGAPNITANLTSSLLSVCGWSQINPEDDTPDGHADLVLYITRFDLELPDGNRQVRG 224	Db	
OY	121 VTQGGACSPWWSCLITDGTGFDLGVTI 148	OY	
Db	225 VTQGGACSPWWSCLITDGTGFDLGVTI 252	Db	
			RESULT 9
			AAB42668
ID	AAB42668 standard; Protein; 118 AA.	ID	
XX		XX	
AC	AAB42668;	AC	
DT		DT	
XX	08-FEB-2001 (first entry)	XX	
XX		XX	
DE	Human ORFX ORF2432 polypeptide sequence SEQ ID NO:4864.	DE	
XX		XX	
KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;	KW	
KW	vulvular; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;	KW	
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;	KW	
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;	KW	
KW	hypotensive; dermatological; immunosuppressive; antitumour;	KW	
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;	KW	
KW	antianemic; gene therapy; cancer; proliferative disorder; hypertension;	KW	
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;	KW	
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;	KW	
KW	cholesterol ester storage; systemic lupus erythematosus; infection;	KW	
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;	KW	
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;	KW	
KW	bone damage; cartilage damage; antinflammatory disease; coagulation;	KW	
KW	thrombosis; contraceptive.	KW	
XX		XX	
OS	Homo sapiens.	OS	
XX		XX	
PN	WO200058473-A2.	PN	
XX		XX	
PD	05-OCT-2000.	PD	
XX		XX	
PF	31-MAR-2000; 2000WO-US08621.	PF	
XX		XX	
PR	31-MAR-1999; 99US-0127607.	PR	
PR	02-APR-1999; 99US-0127636.	PR	
PR	05-APR-1999; 99US-0127728.	PR	
XX	30-MAR-2000; 2000US-0540763.	XX	
PA	(CURA-) CURAGEN CORP.	PA	
XX		XX	
PI	Shankers RA, Leach M;	PI	
XX		XX	
DR	WPI; 2000-602362/57.	DR	
DR	N-PSDB; AAC76877.	DR	

Db	122 VTQGGACSPWWSCLITDGTGFDLGVTI 149	CC	after gastrointestinal irradiation
----	-------------------------------------	----	------------------------------------


```

XX PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX PS Claim 11; Page 4050; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatocrotic; vulnery;
CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
CC osteopachic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX SQ Sequence 118 AA;
Query Match 68.4%; Score 529; DB 21; Length 118;
Best Local Similarity 97.1%; Pred. No. 1.7e-20;
Matches 100; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 46 LGAQFRVHLVWVILTEPEGAPNITANLTSSLVSVCWSQTINPEDTDFGHADLVLYIT 105
DB 1 LGAFRHLVWVILTEPEGAPNITANLTSSLVSVCWSQTINPEDTDFGHADLVLYIT 60
QY 106 RFDELDPGNKVRGVLTQLGACSPWWSCLITDGTGDLGVTI 148
DB 61 RFDELDPGNXAVRGVLTQLGACSPWWSCLITDGTGDLGVTI 103
RESULT 10
AAU05396
ID AAU05396 standard; Protein; 26926 AA.
XX AC AAU05396;
XX
XX DT 24-OCT-2001 (first entry)
XX DE Human titin (connectin) protein sequence.
XX
XX KW Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;
XX KW titin-related disease; zebrafish; heart failure; heart disease.
XX
XX OS Homo sapiens.
XX
XX PN WO200151666-A1.
XX
XX PD 19-JUL-2001.
XX
XX PF 12-JAN-2001; 2001WO-US01212.
XX
XX PR 12-JAN-2000; 2000US-0175787.
XX
XX (GENO ) GEN HOSPITAL CORP.
XX
XX PA Fishman MC;
XX
XX PI
XX
XX DR WPI; 2001-451869/48.
XX N-PSDB; AAS05390.

```

```

XX PT Determining if a subject has or is at risk of developing a
PT titin-related disease or condition, particularly heart failures,
PT comprises detecting the presence of a mutation in the titin gene -
XX
XX PS Disclosure; Page 57-111; 114pp; English.
XX
CC The present sequence representing human titin (also known as connectin)
CC is described in an invention relating to a novel method for determining
CC whether a subject has or is at risk of developing a titin-related
CC disease or condition. The method comprises analysing a nucleic acid of
CC a sample from the subject and detecting the presence of a mutation
CC (e.g. the pickwick mutation in the cardiac specific exon N2B) in the
CC titin gene, which indicates that the subject has or is at risk of
CC developing a titin-related disease. The zebrafish which has a phenotype
CC similar to mammalian heart failure is used as a model. The method is
CC useful for detecting an increased likelihood of heart disease, such as
CC heart failure, in a patient, so that appropriate intervention can be
CC instituted before any symptoms occur. The method may also be used to
CC facilitate determination of etiology of an existing heart condition,
CC such as heart failure, to identify compounds that can be used to treat
CC or prevent heart conditions, in prenatal genetic screening, e.g. to
CC identify parents who may be carriers of a recessive titin mutation.
CC Compounds identified using the methods may be used to treat patients
CC that have or are at risk of developing heart disease, e.g. heart
XX
XX SQ Sequence 26926 AA;
Query Match 35.2%; Score 272; DB 22; Length 26926;
Best Local Similarity 0.4%; Pred. No. 0.45;
Matches 101; Conservative 16; Mismatches 23; Indels 24887; Gaps 28;
QY 9 ELL-----VAQPDV----- 11
DB 190 ELLVQEEVEVPAAKTKTIVSTAQISESRQTRIEKKIEAHFDARSIAIVEMVIDGAAGQOL 249
QY 12 -----VAQPDV----- 11
DB 250 PHKTPPIPPKPSRPTPPSIAAQAARQSPSPIRHPSFVRHVRAPTPSPVRSP 309
QY 12 -----VAQPDV----- 18
DB 310 AARISTPISRSVRLMRKTOASTVATGPEVPPMPKQEGYVASSSEAEEMRETTLTSTQ 369
QY 19 ----- 18
DB 370 INTEWEGRYGVQEQVTISGAAGAAASVSASAYAAEAVATGAKEVKODADKSAAVATV 429
QY 19 ----- 18
DB 430 VAAVDMARVREPVISAVEQTAQRTTTTAVHIQPAQEQVRKEAEKTAVTKVVAADKAKEQ 489
QY 19 ----- 18
DB 490 ELKSRTKEIITTKQEQMHVTHEQIRKETEKTEFPVKYVVISAAKAKEQETRISEBITKKQK 549
QY 19 ----- 18
DB 550 VTQEAIMKETRTVVPVKVIVATPKVKEQDLVSRGREGITTKREQVQITQEKVRKEAEKTA 609
QY 19 ----- 18
DB 610 LSTIAVATAKAKEQETILTRETMTATROEQIQVTHGKVDVGKAEAVATVVAADQAVR 669
QY 19 ----- 18
DB 670 EPREPGHLEESYAQOTTLEYGYKERISAAKVAEPPORPASEPHVVPKAVKPRVIOAPSET 729
QY 19 ----- 18
DB 730 HIKTTDQGMHISQIKKTTDLTTLRLVHVDKRPRTASPHFTVSKISVPKTEHGYEASTA 789

```

QY	19	-----	18
Db	790	GSAIATLQKELSATSAAKITKSVKAPTAKPSETRVABPTPLPOFPADTPTYKSEAG	849
QY	19	-----	18
Db	850	VEVKVCVSITGTTVREERFVLHGREAKVTETARPAPVPIVTPPTLVSLGNKVTVI	909
QY	19	-----FQ-----AHQEDTERVY-	30
Db	910	EGESVTLCHISGYPSTVTWYREDYQIESSIDFOITFQSGIARLMIREAFDSGRFTC	969
QY	31	-----	30
Db	970	SAYNEAGTVSTCVLAVOVSEEBKETTAVTEKFTTEERKRFVESRDVMTDTSLTEOAG	1029
QY	31	-----	30
Db	1030	PGEPAAPYFITKPVVQKLVEGGSVFGCVGGNPKPHVVYKKSQVPLTTGTRYKYVSNKQ	1089
QY	31	-----	30
Db	1090	TGECKLVISMTFADDAAGEYTVVRNKHGETSASASLLEEADYELLKMSQEBMLYQTOVTA	1149
QY	31	-----	30
Db	1150	FVQPEVGETAGPFVYSEYEKEKSQALLRKMAKDTVVVRYTVEDEFFHISSEFERLI	1209
QY	31	-----	30
Db	1210	KEIYRIIKTTLEELLEEDGEKKMAVDISESEAVESGFLRIKNRYILEMGVTFHCQMS	1269
QY	31	-----	30
Db	1270	CYPLPKIANYKDKRIKHGBERYQWDFLQGRASLRIPVVLPEDEGIYTAFASTNKGNAIC	1329
QY	31	-----	30
Db	1330	SGKLYVEPAAPLGAPTYIPTLEPVSRISLSPRSVSRSRPSRMSPARMSPARMSPA	1389
QY	31	-----	30
Db	1390	RMSGRLEETDESQERLYKPVVLKPVSKFCLEGOTARFDLKVGRPMPTETFWHDQG	1449
QY	31	-----	30
Db	1450	QIVNDYTHKVIKEDGTQSLIIVPATPSDGEWTVVAQNAGRSSISVILTVEAVEHQVK	1509
QY	31	-----LTNLNI-----36	
Db	1510	PMFVEKLKNVNIKEGSRLEMKVRATGNPNPDIIVWLKNSDIIIVPHKYPKIRIEGTKGAAL	1569
QY	37	-----	36
Db	1570	KIDSTVSDSAWYTATAINKAGRDTTRCKNVVEFEAFPEPERKLIIPRGTYRAKEIAAP	1629
QY	37	-----	36
Db	1630	ELEPLHLYRQOEWEGBDLYDKEKQKPPFKKLTSLRKRFPAHFECRLTPISDPTMW	1689
QY	37	-----	36
Db	1690	VEWLHDGKPLEAANRLRMINEFCYCSLDYGVAYSRDSGIITCRATNKYGTDTSATLIVK	1749
QY	37	-----GA-----38	
Db	1750	DEKSLVEESQLPEGRKGLORIEELERMAHEGALTGVTTDQEKQKPDIVLYPEPVRVLEG	1809
QY	39	-----	38
Db	1810	ETARFRCRVTGYPOPKNVYLNQGLIRKSRFRVRYDGIHYLDIVDCQSDYDTEGVKVTAE	1869
QY	39	-----	38

Db	1870	NPEGVIEHKV	LEIQOREDFRSVLRAPRPF	FHVHBEKGLQFVQKVDRAVD	TTETKE	1929															
Qy	39	-----	-----	-----	ELLR---	42															
Db	1930	VVKLKRAERI	THEKVPESSELSRKFRRTEG	YIEATAVELSKRKKDSYE	ELLRKT	1989															
Qy	43	-----	-----	-----	-----	42															
Db	1990	DELLHWTKE	TEBEKKALABEGKITI	TPKPKIELSPSMEAPKIF	ERIOSQT	2049															
Qy	43	-----	-----	-----	-----	45															
Db	2050	HPRVVRVCK	PDPECEWYKNGKIERSDRI	YWHPEDNVCELVIRDVTA	EDSASIMVKAIN	2108															
Qy	46	-----	-----	-----	-----	45															
Db	2110	IAGETS	SHAFLLVQAKQLIT	FTOELQDVVAKEDT	MATFECETSEPFVKVWKYK	DMGMEVH	2169														
Qy	46	-----	-----	-----	-----	45															
Db	2170	EGDKYRM	HSRDKVFLSIL	ITDTSDAEDYSCVLVEDEN	VKTAKLI	VEGAVVEFKELQD	2229														
Qy	46	-----	-----	-----	-----	45															
Db	2230	IEVPESY	GELCIVSPENIEGKYHND	VELKSNGKYITISRRGRQ	NLTVKDVTKED	QGE	2289														
Qy	46	-----	-----	-----	-----	45															
Db	2290	YSFVIDG	KTTCKLKMKPRPIA	ILOGLSQKVCED	IVOLEVKVLS	VEGVMKDGQEV	2349														
Qy	46	-----	-----	-----	-----	45															
Db	2350	QPSRVH	IVIDKQSHMLLI	EDMTKEDAGNYSFTI	PALGLSTSGRVSVY	SDVITPLKDVN	2409														
Qy	46	-----	-----	-----	-----	45															
Db	2410	VIEGKAV	LECKVSPDVT	SVKWLND	EQIKPDDRQV	AIKGTQKQLVIN	RTHASDEGPY	2469													
Qy	46	-----	-----	-----	-----	-----	45														
Db	2470	KLIVGRV	ETNCNLSVEKI	IIIRGLDLTCTETQ	NVVFEVLSHSGID	VLWNFKDK	EIKPS	2529													
Qy	46	-----	-----	-----	-----	-----	45														
Db	2530	SKYKIEA	HGKIYKLTVLN	NMKDEGKYTFY	AGENMTSGKLT	VAGGAI	SKPLTDQTVAESQ	2589													
Qy	46	-----	-----	-----	-----	-----	45														
Db	2590	EAVFECE	VANPDSKG	BWRDGHKLPL	TNNIR	ESDGHKRLII	IAATKLDDEGEY	TYKVAT	2649												
Qy	46	-----	-----	-----	-----	-----	45														
Db	2650	SKTSAKL	QVEAVKIKKTL	KNLTVTETQDA	VFTVELTHPNV	KGQVQWIKNGV	VVLESNEKYAI	2709													
Qy	46	-----	-----	-----	-----	-----	58														
Db	2710	SVKGTI	YLSRIKNC	AI	VD	ESVYGFRLGR	LGASARLHVETKII	KKPKDVALENAT	VAPE	2769											
Qy	59	-----	-----	-----	-----	-----	58														
Db	2770	VSVSHD	TVPVKWFHKS	VEIKPSDKHRL	YSEKVKHKL	MLONIS	PSDAGEY	TA	VVGOLECKA	2829											
Qy	59	-----	-----	-----	-----	-----	58														
Db	2830	KL	FETLHITK	MNI	EV	PETK	TASFECEVSHFN	VP	SMWLNKNGVEI	EMSEBKFI	VVOGKL	2889									
Qy	59	-----	-----	-----	-----	-----	58														
Db	2890	HQLI	IMNT	ST	ED	SAEYTF	VC	GN	DQ	VSATLTV	TP	MIT	SMLKD	IN	ABEK	OTI	TE	FT	V	NYE	2949
Qy	59	-----	-----	-----	-----	-----	58														

Db 2950 GISYKWLKNGEIKSTDKCOMRTKTLTHSLNIRNVHFGDAADYTFVAGKATSTATLYVEA 3009
QY 59 ----- 58
Db 3010 RHIEFRKHIDIKVLEKKRAMFECEVSEPDITVOMMKDOELOITDRIKIQEKYVHRL 3069
QY 59 ----- 58
Db 3070 IPSTRMSDAGKYTVVAGGNVSTAKLFVEGRDVRIRSIKKEVQVIEKORAVVEFVNEDDV 3129
QY 59 ----- 58
Db 3130 DAHWYKDGIEINQOERHKYVVERIRHRMFISETROSDAGEYTFVAGRNRSSVTLYVNA 3189
QY 59 ----- 58
Db 3190 PEPPQVLQELQVTVQSGKPARFCAMISGRPOPKISWYKEQLLSTGFKCKFLHDQOBYT 3249
QY 59 ----- 58
Db 3250 LLLIEAFPEDAANYTCEAKNDYGVAATTSASLSVEVPEVSPDQEMPVYPPIITPLQDTV 3309
QY 59 ----- 58
Db 3310 TSEQPARFCRVSGTDLKVSWSYKDKKIKPSRFRMTQFEDTYQLEIAEAYPEDEGYT 3369
QY 59 ----- 58
Db 3370 FVANNAYQVSSSTANLSLEAPESILHERIEOIEEMEMKEPSSSFLSAEEGLHSAEQLS 3429
QY 59 ----- 58
Db 3430 KINETLELLESPPYPTKFDSEKGTGPIFIKEVSNADISMGDVATLSVTVIGIPKPKIQ 3489
QY 59 ----- 58
Db 3490 WFNGLVLTSPADYKFVFGDDHSLIILFTKLEDEGEYTCMASNDYGTICSAYLKINSK 3549
QY 59 ----- 58
Db 3550 GEGHKOTETESAVAKSLEKLGCGPCPPHFLKELPIRCAQGLPAIFEYTVVGEPAPTWTF 3609
QY 59 ----- 58
Db 3610 KENQOLCTSVYTIHNPNGSGTFIVNDPOREDSGLYICKAENMLGESTCAEELLVLED 3669
QY 59 ----- 58
Db 3670 TDMTDTPCAKASTPEAPEDFPQTLKGPVAVELDSEOEIATFVKDTILKAALITEENQOL 3729
QY 59 ----- 58
Db 3730 SYEHIKANELSSOLPLGAQELQSILEODKLTPESTREFLCINGSIHFOPLKEPSNLQL 3789
QY 59 ----- ILTEPEG- 65
Db 3790 QIVOSQKTFSEKILMPEEPETQAVLSDEKIFPSMSIEQINSITVEPLKTLAAEPGN 3849
QY 66 ----- 65
Db 3850 YPOSSIEPPMHSYLTSAEVEVLSLKEKTVSDTNREQVTLQKQEAQSALILSQAEGHV 3909
QY 66 ----- 65
Db 3910 ESLOSPOVNI SOVNYEPLVPSEHSCTEGGKILIESANPLENAGQDSAVRIEKGSLRPPL 3969
QY 66 ----- 65
Db 3970 ALEEKQVLLKEHSDNVMPDPIIESKREPVAIKKVOEGRDLLSKESLLSGIPEQR 4029
QY 66 ----- 65
Db 4030 LNLKIQICRALQAAVASEQPLGFSEWLRNIEKVEEAVNITQEPHRHIMCMVLTSAKSVT 4089

QY 66 ----- 65
Db 4090 EEVTTIIIEDVDQMANLKMELRDALCAIIEIDILTAEPRIQOAKTSLQEEMDSFG 4149
QY 66 ----- 65
Db 4150 SOKVEPITEPEVESKYLJISTEEVSFNVQSRVKYLDATPVTKGVASAVVSDKQDES LKP 4209
QY 66 ----- 65
Db 4210 SEEKESSSESCTEEVATVKIOEAEGLIKEDGPMIHTPLVDTVSEEGDIVHLTTSITNA 4269
QY 66 ----- 65
Db 4270 KEVNYFENKLVPSDEKFKCLQDQNTYTLVDKVNTEHOGVEYVCEALNDSKGTATSAKL 4329
QY 66 ----- ANI- 69
Db 4330 TVVKRAAPVIKRKIBPLEVALGHLAKFTCEIOSAPNVRQWFKAGREIYESDKSIRSK 4389
QY 70 ----- 69
Db 4390 YISSLEILRTQVDCGEYTCKASNEYSVCTATLTVTVPGEKKVRKLLPERKPEKEE 4449
QY 70 ----- 69
Db 4450 VVLKSVLRKRPEEBEPKVEPKLEKVKPAPVPEPPPKPEVEEVPVTVTKRERKIPEPTK 4509
QY 70 ----- 69
Db 4510 VPEIKPAIPAPKPKPEAEVTKIKPPPVEPEPTPIAAPVTVPVGKKAERKPEEA 4569
QY 70 ----- 69
Db 4570 AKPGPIKGVPKTSPPIEAERKLRPGSGEKPDEAPFTYQLKAVPLKFVKEIKDIL 4629
QY 70 ----- 69
Db 4630 TESEFVGSSAIFECVSPSTAITTWMKGSNIRESPKHFIA DGKORKLHIIDVQLSDAG 4689
QY 70 ----- 69
Db 4690 EYTCVLRNGNEKTSTAKLVVEELPVRFVKTLBEEVTVVKGQPLYLSCELNKERDVVWRK 4749
QY 70 ----- 69
Db 4750 DGKIVVEKGRIVPGVIGLMRLALTINDADDTAGTYTVTVNANNLECSCKVKEVIRD 4809
QY 70 ----- 69
Db 4810 WLKPIRDQHVKPKGTAFACDIAKDTPNIKWPKGYDEIPAEPNDKTEILRDGNHLYLKI 4869
QY 70 ----- 69
Db 4870 KNAMPEDIAEYAVEIEGKRYPAKJLGEREVELLKPIEDVTIYEKESASPDABEISADIP 4929
QY 70 ----- 69
Db 4930 GQWKLKCELLRPSPTCEIKAEKGKRFILHLKVKLDOAGEVLYQALNAITTAILTVEIEI 4989
QY 70 ----- 69
Db 4990 DFAVPLKDVTPERRQARPECVLTREANVTWSKGPDIKIKSSDKFDIADGKHLVINDS 5049
QY 70 ----- 69
Db 5050 QFDEGVYTAEBEGKKTSARLFVTGIRLKFMSPLEDQTVKEGTATFVCELSHEKHHVW 5109
QY 70 ----- 69
Db 5110 FKNDAKLHTSRTVLIISSEKTHKLEMKEVTLDISQIAQVKELSSSTAQLKVLADPYFT 5169

QY	70	-----	69
Db	5170	VKLHDKTAVEKEITLKEVSKDVPVKWFKDGEIIVPSKYSIKADGLRILKIKKADLK	5229
QY	70	-----	69
Db	5230	DKGEYVDCDGTDKANKVTVEARLIEVEKPLYGVEVGETAHFEIELSEDPVHGOWKLK	5289
QY	70	-----	69
Db	5290	GQPLTASPDCIEIEDGKGHILILHNCQLGTMTEVSVFOAANAKSAANLKVLEPLIFITPL	5349
QY	70	-----	69
Db	5350	SDVKVFEKAEKPECEVSRPKTFRWLKGTQEIITGDDRFELIKDGTKHSMVKSAAPEDE	5409
QY	70	-----	69
Db	5410	AKYMFEBDKHTSGKLIIEGIRLKLPLKDVTAKEKESAVTVELSHDNIRVKWFKNDQ	5469
QY	70	-----	69
Db	5470	RLHTRSVSMODEGKTHSITFKDLSIDDTSQIRVENMGMSSEAKLTVLEGDPYFTGKLQD	5529
QY	70	-----	69
Db	5530	YTGVEKDEVILQCEISKADAPVKWFKDGEIKPSKNAVITGDKKRMILKALKSDIGQ	5589
QY	70	-----	69
Db	5590	YTCDCGDKTSKGLDIEDREIKLVRPLHSVEMETETARFETEISEDDIHANWKLKGEAL	5649
QY	70	-----	69
Db	5650	LOTPDCEIKECKIHSLVHLNCRLOQTCGVDFQAAVKKSSAHLRVKPRVIGLLRPLKDV	5709
QY	70	-----	69
Db	5710	VTAGETATDCELSYEDIPVEWYLGKLEPSDKVPRSEKGVHTLTLRDVKLEDAGEVQ	5769
QY	70	-----	69
Db	5770	LTAQFKTHANLKVKEPPVEFTKPLEDQTVREGATAVLECEVSRENAKWFKNGTEILK	5829
QY	70	-----	69
Db	5830	SKKYEIVADGRVKLVIHDCPTEDIKTYTCDAKFKTSCNLNVPPHVEFLRPLTLQVR	5889
QY	70	-----	69
Db	5890	EKEMARFECESRENAKVKFKDGAIEIKGKYDIISKGAVRILVINKCLLDDEAEYSCE	5949
QY	70	-----	69
Db	5950	VRTARTSGMLTVLEEAFTKNLANIEVSETDTIKLVCEVSKPGAIEVIWYKGDEEIIETG	6009
QY	70	-----	69
Db	6010	RYEILTEGRKRLVIONAHLEDAGNYNCRPLSSRTDQVKVHVAEAFISKPNLEILEG	6069
QY	70	-----	69
Db	6070	EKAEFVCSISKESFPQWKRDDKTLESQKYDVADGKKRVLVVKDATLQDMCTYVVMVG	6129
QY	70	-----	69
Db	6130	AARAAHLTVIEKLRIVVPLKTRVKEQEVWFNCEVNTTEGAKAKWFRNEEAFDSSKYI	6189
QY	70	-----	69
Db	6190	ILQKDLVTLIRDAHLDDQANVNSLTNHRGENVKSAANLIVEEEDLRIVEPLKDIETW	6249
QY	70	-----	69

Db	6250	EKSVTFWCKVNLNLVTLKWTKNGBEVPFDRVSVYRVDKYKHLMTIKOCPPDEGEYIVT	6309
QY	70	-----	69
Db	6310	AGQDKSVAELLIIEAPTEFVEHLEDTVTEPDDAVFSQLSREKANVKYRNGREIKEGK	6369
QY	70	-----	69
Db	6370	KYKFEKDSIHRLIILKDCRLDDECEVACGVEDKRSARLFBEEIIRPPODILEAPG	6429
QY	70	-----	69
Db	6430	ADVFLAELNKDKVEQWLRNNWVVGDKHOMSEGIHRLQICDIKPRDOGEYRFAIK	6489
QY	70	-----	69
Db	6490	DKEARAKLELAAPKIKTADQDLVVDVGKPLTMVVPYDAYPKAAEAWFKENEPLSTKTID	6549
QY	70	-----	69
Db	6550	TTAEQTSFRILEAKGDKGRYKIVLQNKHGAEGFINLKVIVDPGVRNLEVTETPDGEV	6609
QY	70	-----	69
Db	6610	SLAWBEPLTDGGSKIIGYVYVVERRDIKRKTWVLATDRAESCEPTVTGLQGGVEYLFVSA	6669
QY	70	-----	69
Db	6670	RNRVGTGEVETNPVEARSKYDVPGPLNVITIDVNRFGVSLTWEPPEYDGAETNYV	6729
QY	70	-----	69
Db	6730	IELRDKTSIRWDTAMTVRAEDLSATVTVVGOEYSFRVRAQNRIQVGPASAATPFVKVA	6789
QY	70	-----	76
Db	6790	DPIERPSPVNLTSSTQTSVOLKWEPPUKOGGSPILGVIIERCCEGKNWIRCNKLV	6849
QY	77	-----	76
Db	6850	PELTYKVTGLEKNKYLRYVSAENKAGVSDPSEILGPLTADDAFVEPTMOLSAFKDGLV	6909
QY	77	-----	76
Db	6910	IVPNPITILVPSTGYPRPTATWCFGDKVLETGDRVKMKTLSAYAEVLVSPSERSDKGIYT	6969
QY	77	-----	76
Db	6970	LKLENRVKTTISGEIDVNVVIARPSAPKELKFGDITKDSVHLTWBPDDDDGSPLTGYVVEK	7029
QY	77	-----	81
Db	7030	REVSRTWTKVMDPFTDLEFTVPDLVQGEYLFVKVCARNKCGPGEPAYVDEPVNMSTPAT	7089
QY	82	-----	81
Db	7090	VPDPENVKWRDRTANSIFLTWDPKNDGGSRIKGYIVERCPGSDKXWVACGEPVAETKM	7149
QY	82	-----	81
Db	7150	EVTGLEGGKWAYRVKTLNROGASKPSRPTEEIOAVDTQEAPEIFLDVKLLAGLTVKAGT	7209
QY	82	-----	81
Db	7210	KIELPATVTKPEPKITWTAKMILKQDKRITENVPKSTVTIVDSKRSDTGTVIIIEAV	7269
QY	82	-----	81
Db	7270	NVCGRATAVVEVNLDPGPPAADIPTVNESCLLTWNPPRDGGSKITNYVVERRATD	7329
QY	82	-----	81

Db 7330 SEVHKLSSVTKDNFKATKLIIPNKEYIFRVAENMYGAGEPVQASPIITAKYQFDPGPP 7389
Qy 82 ----- 81
Db 7390 TRLEPSDITKDAVTLTWCEPDDGGSPITGYWVERLPDPTDKWRCNKMVPVKDITTYRVKG 7449
Qy 82 ----- 81
Db 7450 LTNKKYFRVLAENLAGPKSPKSTEPILIKDIPDPWPGKPTVKDVCKTSVRLNWK 7509
Qy 82 ----- 81
Db 7510 PEHGGAKIESYIEMLKTGTDEMURVAEGVPTTQHLLPGLMEQEYSFRVRAVNKAGES 7569
Qy 82 ----- 81
Db 7570 EPSEPSDVLCREKLYPPSPRMLEVINITKNTADLKWTPKXDKGSPITNYIVEKRDVR 7629
Qy 82 ----- 81
Db 7630 RKGWQTVDTTVTKCTVTPLTGSLYVFRVAENAIQSDYTEIEDSVLAKDTFTTGP 7689
Qy 82 ----- 81
Db 7690 PYALAVVDVTKRHVDLKWEPKNDGGRPIQRYVIEKKERLCTRWWKAGTAGDCNFRVT 7749
Qy 82 ----- 81
Db 7750 DVEGTEVOFQVRAENAGVHPSEPTEILSIEDPTSPSPPLDLHVDHVDAGRKHIAIAWK 7809
Qy 82 ----- 81
Db 7810 PPEKNGSGPIGYHVEMCPVCTEKMVRNRPDKLKFVEGVVPDKKEYLVRVAVNAI 7869
Qy 82 ----- 81
Db 7870 GVSEPSEISNVAKDPCKPTIDLETHDIIIEGEXLSIPVPPRAVPVPTVSMHKDGE 7929
Qy 82 ----- 81
Db 7930 VKASDRLTMKNDHISAHLEVPKSVRADAGIYITLENKLSATASINVKVLGPCKDI 7989
Qy 82 ----- 81
Db 7990 KASDITKSSCKLTWEPFDDGTPILHVLERERAGRTYIPVMSGENKLSWTVKDLIPN 8049
Qy 82 ----- 81
Db 8050 GEYFRVKAIVKVGGEYIELKNPVIAQDPKQPPDVPDVEVHNPTAEAMITWKPLYD 8109
Qy 82 ----- 81
Db 8110 GSKIMGYIIEKIAKEERWRCNEHLVILTITAKGLEEGKEYQFRVRAENAGISEPS 8169
Qy 82 ----- 81
Db 8170 RATPPTKAVDPIADPKVILRTSLEVYKRGDEIALDASISGSPYPTITWIKDENWIVPEIK 8229
Qy 82 ----- 81
Db 8230 KRAAPLVRRRKGEVQEEBPFVPLTQRLSIDNSKKGESQLRVDRSLRDPHGLYMIKVEND 8289
Qy 82 ----- 81
Db 8290 HGIAPACTVSVDTPGPPINPFVFDIRKTSVLCKWEPLDDGGSEIINYTLEKDKTKP 8349
Qy 82 ----- 81
Db 8350 DSEWIVVTSTLRHKYSVTKLIEGKEYLFRVRAENRFGPGPCVSKPLVAKDPFGPPDAP 8409
Qy 82 ----- 81
Db 8410 DKPIVEDVTSNMLVKWNEPKDNGSPILGYLEKREVNSTHWSRVNKSLLNALKANVDGL 8469

Qy 82 ----- 81
Db 8470 LEGLTYVFRVCAENAAAGPKFSPSPDKTAHDPISPPGPIPRVTDTSSTTTELEWEPPA 8529
Qy 82 ----- 81
Db 8530 FNGGGEIVGYFVDKOLVGTNKNWRCRTEKMKVQVTVKEIREGADYKLRVSAVNAAGEOP 8589
Qy 82 ----- 81
Db 8590 PGETQVTVAEQEPPEPAVELDVSKGGIOIMAGKTLRIPAVVTGRPVTPKVTKEEGELD 8649
Qy 82 ----- 81
Db 8650 KDRVVIDNVGTSKELIIKDALRKHGRYVITATNSCGSKFAAAARVEFVDPGVPVLDLKP 8709
Qy 82 ----- 81
Db 8710 VTNRMCLLNWSDPEDDGGSEITGFIERKDAKMTWROPIETERSKCDITGLLEGOEYK 8769
Qy 82 ----- 81
Db 8770 FRVIAKNKFCGCPVEIGPILAVDPLGPPTSPERLTYTTERQSTITLDMKEPRSNNGSPI 8829
Qy 82 ----- 81
Db 8830 QGVIIERRHRDKDPFERNVKNRLCPTTSFLVENLDEHOMYEFVRKAVNEIGESEPSLPLNV 8889
Qy 82 ----- 81
Db 8890 VIQDDEVPTIKLRSLVRGDTIKVKAGEPVHIPADVTGLPMPKIEWSKNETVIEKPTDAL 8949
Qy 82 ----- 81
Db 8950 QITKEEVSERSEAKTELSIPKAVREDKGTYYVTASNRLGVSFRNVHVEVYDRPSPRNLAV 9009
Qy 82 ----- 81
Db 9010 TDIKABSCYLTWDAPLDNGGSEITHYVIDKRDASRKAEEVNTAVKRYGIWKLIPN 9069
Qy 82 ----- 81
Db 9070 GOYEFVRVAVNKYISDECKSDKVIQDPYRLPGPGPKVYLARTKGSMLVSWTPPLDNG 9129
Qy 82 ----- 81
Db 9130 GSPITGYWLEKREEGSPYMSRVSRAPITKVGLKGVEFNVPRLLEGVKYOFRAMAINAAGI 9189
Qy 82 ----- 83
Db 9190 GPPSEPSDPEVAGDPPIFPPOPPSCPEVKDKTKSSI SLGWMKPPAKDGGSPIKGYIVEMQEE 9249
Qy 84 ----- 83
Db 9250 GTTDMKRVNEPDKLIITTCCECVNPNLKLRYFRVKAIVNEAGESEPSDITGTEIPATDIOE 9309
Qy 84 ----- 83
Db 9310 EPEVFDIGAQCCLVCKAGSOIRIPAVIKGRPTPKSSWEFDGKAKAMKDCVHDIPEDAQ 9369
Qy 84 ----- 83
Db 9370 LETAENSSVIIIECKRSHTGKYSITAKNKAQOKTANCRVKVMDVPGPKDLKVSDITRG 9429
Qy 84 ----- 83
Db 9430 SCRLSWKXWPDGDDGDRIKGYVIEKRTIDGKAWTKVNPDCGTTFFVVPDLLSEOOYFFRVR 9489
Qy 84 ----- 83
Db 9490 AENRFGIGPPVETIORTTARDPIYPPDPPIKLIKGLITKNTVHLSMKPKPKNDGGSPVTHY 9549

Db 11710 LSGVLTVKAGDTIRLEAGVRGKPFPEVAWTKDKDADTLTRSPRVKIDTRADSSKFSLTKA 11769
Qy 102 ----- 101
Db 11770 KRSDGGKYVVTATNTAGSFVAYATVNVLDKPGPVRLKIVDVSSDRCTVCWDPEDDGGC 11829
Qy 102 ----- 101
Db 11830 EIQNYILEKETKRMVWSTYSATVLTGCTTVTRLIEGNEYIFRVRAENKIGTGPTTESKP 11889
Qy 102 ----- 101
Db 11890 VIAKTYDKPCRDPPEVTKVSKEMTVMNPPEYDGGKSITGYFLEKKEKHSTRWVPVN 11949
Qy 102 ----- 101
Db 11950 KSAIPERRMKVQNLPLDHEYQFRVKAENEIGIGESPLSRPVPVAKDPIEPGPPTNFRV 12009
Qy 102 ----- 101
Db 12010 DTTKHSITLGWGPYDGGAPIIGYVVMRPKIADASPDEGMRKCNAAAQLVRKEFTVTS 12069
Qy 102 ----- 101
Db 12070 LDENQVEFRVCAQNOVGIGRPAELKEAIKPKILEPPEIDLDASMRKLVIVRAGCPRL 12129
Qy 102 ----- 101
Db 12130 FAIVRGVPAPKVTWRKVGIDNVVRKGQVDLVTWAFVIPNSTRDDSKYSLTLVNPAGE 12189
Qy 102 ----- 101
Db 12190 KAVFVNRVLDTGPGVSDLKVSDVTKTSCHVSWAPPENDGGSQVTHYIVEXREADRKTWS 12249
Qy 102 ----- 101
Db 12250 TVTPEVKKTSFHVTLNPGNEYFRVTAVNEYGPGVTPDKPVLASDPLSEDPPRKLE 12309
Qy 102 ----- 101
Db 12310 ATEMTKNSATLAWPLPRDGGAKIDGYIISYREEQPADRWTEYSVVKDLSLVVTGLKEG 12369
Qy 102 ----- 101
Db 12370 KKYFRVAARNAGVSLPREAGVYEAKQLLPKILMPEQITIKAGKKLRIEAHVYCKP 12429
Qy 102 ----- 101
Db 12430 HPTCKWKGEDEVTSLSLAVHKADSSILIIKDVTRKDSGYISLTAENSSGTDQKIKV 12489
Qy 102 ----- 101
Db 12490 VMDAPGPPQPFDISDIDADACSLSWHIPLEDDGSNTNIVIEKCDVSRGDWVTALASV 12549
Qy 102 ----- 101
Db 12550 TKTSCRVGKLIPOQEYIFRVRAENRFGISEPLTSPKMWAPFPFGVSEPKNARVTKVNDK 12609
Qy 102 ----- 101
Db 12610 CIFVADRDPDGGSPIIIGYLIERKERNLLWVKANDTLVRSTEYPCAGLVEGLEYSFRI 12669
Qy 102 ----- 101
Db 12670 YALNAGSPPSPKTEYVTARMVPDPPGKPEVIDVTKTVSLIWARPKHDKGSKIIIGFV 12729
Qy 102 ----- 101
Db 12730 EACKLPDQKWRNCTAPHQI POEYATATGLBEKAQYQFRAIARTAVNISPPSESDPVTI 12789
Qy 102 ----- 101
Db 12790 LAENVPPRIDLSVAMKSLTLVRAGTNVCLDATVFGKMPMTVSWKKDGTLLKPAEGIKMAM 12849

Qy 102 ----- 101
Db 12850 QRNLCLELFSVNRKDSGDYTTITAENSSGKSATIKLKVLDKPGPPASVKINKMYSDRAM 12909
Qy 102 ----- 101
Db 12910 LSWEPPLEDGGSEITWYIVDKRETSRPNMAQVSATVPITSCSVEKLIIEGHEYOFRICAEN 12969
Qy 102 ----- 101
Db 12970 KYGVGDPVFTEPAIAKPNPDPPGRCDDPPVISNITKDHMTVSMKPPADDGSGSPITGVLLEK 13029
Qy 102 ----- 101
Db 13030 RETQAVNMTKVRKPIIERTLKATGLQEGTEYEFVRTAINKAGPGKPSDASKAAYARDPO 13089
Qy 102 ----- 101
Db 13090 YPPAPPAPKYYDTTRSSVLSWGPAYDGGSPIIGYLVVEVKRADSDNVRCNLQNLQK 13149
Qy 102 ----- 101
Db 13150 TRFEVTLMEDTOYQFRVAVANKIGYSDPSDVPDKHYPKDILIPPEGEHADLRKTLILR 13209
Qy 102 ----- 101
Db 13210 AGVTMLRYVPVKGPPPKITWSKPNVNLDRIGLDIKSTDFTFLRCENVKNYDAGKIYL 13269
Qy 102 ----- 101
Db 13270 TLENSCGKKEYTIWVKVLDTPGPPINVTVEISKDSAYVTWEPPIDGGSPINIYVVKR 13329
Qy 102 ----- 101
Db 13330 DAERKSHSTVTTECSKTSFRVPNLEBCKSYFFRVFAENEYIGDPCGETRADKASQTPGP 13389
Qy 102 ----- 101
Db 13390 VDLKVRSVKSCSIGWKKPHSDGGSRIIGYVVVDFLTEENKQVRVMKSLSLQYSAKDLT 13449
Qy 102 ----- 101
Db 13450 EGKEYTFRVSAENENGEGTPEITVVARDDVAPDLDLGLPOLCYLAKENSFRLKIP 13509
Qy 102 ----- 101
Db 13510 KGKPAPSVSMKGEDPLATDTRVSVESSAVNTTLIVYDCQKSDAGKYITILKNVAGTKEG 13569
Qy 102 ----- 101
Db 13570 TISIKVVGKPGIPTGPIKFDEVTAEAMTLKWAPPKDDGSEITNVIIEKRDSVNNKWVTC 13629
Qy 102 ----- 101
Db 13630 ASAVOKTTFRTRLHEGMEYTFRVSANENYGVGEGKLKSEPIVARHPFDVPDAPPNNIVD 13689
Qy 102 ----- 101
Db 13690 VRHDSVSLTWTDPKKTGSPITGYHLEBFKERNLLWKANKTPIMRDFKVTGLTEGLE 13749
Qy 102 ----- 101
Db 13750 EFRVMAINLAGVGKPSLPSEPVVALDPIDPPGKPEVINITRNSVTLIWTEPKYDGGHKL 13809
Qy 102 ----- 101
Db 13810 GYIIVEKRDLPKSNKMHNVNVPACFTVTDLVEGGKYEFIRAKNTAGAISAPSESTET 13869
Qy 102 ----- 101
Db 13870 IICKDEYEAPTIVLDPTIKDGLTIKAGDTIVLNAISILGKPLPKSSWSKAGDIRPSDIT 13929

Qy	102	-----	101
Ds	13930	QITSTPSSMLTIKVATRKAAGEYITATNPGTKVEHVKVTVLDVPGPPGVEISNVA	13989
Qy	102	-----	101
Ds	13990	EKATLTWTPPLEDGGSPIKSYILEKRETSRLLTWTVSSEDIQSCRHVATKLIQONEVYIFRV	14049
Qy	102	-----	101
Ds	14050	SAVNHKGEVQSEPVKQVDRPGPPGPEKPEVSNVTNTATVSMKRPVDDGGSEITCY	14109
Qy	102	-----	101
Ds	14110	HVERREKSLRWVRAIKTPVSDJCKVTGLQEGSTYEFVSAENRAGIGPPSEASDVLM	14169
Qy	102	-----	101
Ds	14170	KDAAYPPGPPSNPHVTDTTKKSASLAWKPHYDGGLEITGYVVEHOKVGDEAWIKDTTGT	14229
Qy	102	-LVITRF-----	107
Ds	14230	ALRITQVFPDLQTKENYFRISAINDAGVGEPAVIPDVEIVEREMAPDFELDAELRRTL	14289
Qy	108	-----	107
Ds	14290	VVRAGLSIRIFVPIKGRPAPEVTTKDNINLKNRANIENTESFTLLIIPECNRYDTGKV	14349
Qy	108	-----	107
Ds	14350	MTIENPAGKKSGFVNVRLDTPGVNLNRPDTIKDSVTLHWDLPLLDGSRITNYVEK	14409
Qy	108	-----	107
Ds	14410	REATRKYSATTCKCHKTYKVTGLSEGCYFFRVMAENEYIGIETETPEPVKASEAPS	14469
Qy	108	-----	107
Ds	14470	PPDSLNIMDTKSTVSLAWPKPKHGGSKITGYVIEAQRKSDQWTHITTVKGLECVRN	14529
Qy	108	-----	107
Ds	14530	LTEGEYTFQMAVNSAGRSAPRESRPVIVKEQTMPELDRGIYQKLVIAKAGDNKVE	14589
Qy	108	-----	107
Ds	14590	IPVLGRPKPTVMKKDQILKQTORVNFATTATSTILNINECVSRSDSGPYPLTARNIVE	14649
Qy	108	-----	107
Ds	14650	VGDVITIQVHDIPGPPTGPIKFDEVSDFVTFSDPPENDGGVPISNYVEMRQDSTTW	14709
Qy	108	-----	107
Ds	14710	VELATTVIRTYKATRLTTGLEYOFRVKAQNRVCGPGITSAMIVANYPFKVPGPCTPQ	14769
Qy	108	-----	107
Ds	14770	VTAVTKDSMTISWHEPLSDGGSPILGYHVERKERNILQWTVSKALVPGNIFKSSGLTDG	14829
Qy	108	-----	107
Ds	14830	IAYEFVIAENMAGSKSPSEPMALDPIDPPGKPVPLNITRHTVTLKWAKPEYTCGF	14889
Qy	108	-----	107
Ds	14890	KITSYIVKRDLPNGRWLKNFNSILENEFTVSGLTEDAAEFVIAKNAAGAISSPSEP	14949
Qy	108	-----	107
Ds	14950	SDAITCRDDVEAPKIKVDVKFDTVILKAGEAFRLADVSRPPTMEWSKDGELEOTA	15009
Qy	108	-----	107

Ds	15010	KLEIKIADPSTNLVKNKSTRDSCGYTLTATNPGGFAKHIFNVKVLDRPGPEGLAVTE	15069
Qy	108	-----	107
Ds	15070	VTSEKCVLSWPPPLDDGGAKIDHYIVOKRETSRLANTNVASEVQVTKLVTKLKGNEYI	15129
Qy	108	-----	107
Ds	15130	FRVMAVNYGGEPESEPEVLAVNPYPPDPKPKPEVTTITKDSMVVCGHPDSDGSGSEI	15189
Qy	108	-----	107
Ds	15190	INYIVERDKAGORWIKCNKKTLDLRYKVSGLTEGHEYEPRIMAENAGIASPSTPF	15249
Qy	108	-----	107
Ds	15250	YKACDTPFKPPGPNRVLDTSRSSISIAMNKPIYDGGSEITGYMVEIALPEDEWQIVT	15309
Qy	108	-----	107
Ds	15310	PPAGLKATSYTITGLTENQOEYKIRIYAMNSEGLGEPALVPCTPKAEDRMLPEIELDADL	15369
Qy	108	-----	107
Ds	15370	RKVVTIRACCTLRLFVPIKGRPDPEVKWARDHGESLDKASIESASSYTLIVGNVNRFD	15429
Qy	108	-----	107
Ds	15430	GKYILTVNSSGSKSAFVNVRLDTPGPPQDLKVKEVTKTSVTLTWDPPLDGGSKIKNY	15489
Qy	108	-----	107
Ds	15490	IVEKRESTRKAYSTVATNCHTKSMKVDQLQEGCSYFVRLAENEYIGLPAETAESVKAS	15549
Qy	108	-----	107
Ds	15550	ERPLPGKITLMDVTRNSVLSWEKPEHDGGSRIILGYIVEMQTKGSKWATCATVKVTEA	15609
Qy	108	-----	107
Ds	15610	TITGLIQEEYSFRVSAONEKGISDPROLSUPVIAKDLVIPPAPKLLFNTFTVLAGEDLK	15669
Qy	108	-----	107
Ds	15670	VDVPFGRPTPAVTHKDNVPLKQTRVNAESTENNSLLTIKDACREDVGHVVKLTNSA	15729
Qy	108	-----	107
Ds	15730	GEAIETLNVILDKPGPTGPVKMDEVTAADITLSWGPPKPDGSSINNNVEKRDJSTT	15789
Qy	108	-----	107
Ds	15790	TWOIVSATVARTTIKACRLKTGCEYQFIAAENRYCKSTYLNSEPTVAQYPFKVPGPPT	15849
Qy	108	-----	107
Ds	15850	PVVTLSRSDSMEVQNNPEISDGSRGVYHLERKERNISILWVKLNKTPIPOTKFKTTGLE	15909
Qy	108	-----	107
Ds	15910	EGVEYEFVSAENIVGIGKPSKVSECVVARDPCDPPRPEAIIVTRNSVTLQWKKPTYDG	15969
Qy	108	-----	107
Ds	15970	GSKITGYIVEKKELPEGRWMAKASFTNIIDTHPEVTGLVEDHRYEFVRVIAARNAAGVFSEPS	16029
Qy	108	-----	107
Ds	16030	ESTGALTARDEVDPRIKMDPKYKDTIVVHAGESFKVDADYKPIPTIQWIKGDOELSN	16089
Qy	108	-----	107

Db 16090 TARLEIKSTDFATSLSVKDAVRVDSGNVILKAKNVAGERSVTNVKVLDRPPEGPVWI 16149
Qy 108 ----- 107
Db 16150 SGVTAECTLAWKPPLODGSIIINYI VERRETSLRVVVDANVOTLSCKVTKLEBNE 16209
Qy 108 ----- 107
Db 16210 YTFRIMAVNKYGVGELESEPVAKNPVVPDAPKAPVTTVTKDSMIVVHERPASDGS 16269
Qy 108 ----- 107
Db 16270 EILGVLEKRDKEGIRWTRCHKRLIGELURLAVTGLIENHDEFRVSAENAGLSEPPS 16329
Qy 108 ----- 107
Db 16330 AYQACDPIYKPGPNPKVIDITRSSVFLSWSKPIYDGCIEIOGYIVEKCDNVNGWTM 16389
Qy 108 -----DL 109
Db 16390 CTPPTGINKTNIIEKLEKHEYNFRICAINKAGVGEHADVPGPPIIVEEKLEAPDIDL 16449
Qy 110 EL----- 111
Db 16450 ELRKIINIRAGSLRFPVPIKGRPTPEVKWGVKVGDEIRDAAIIDVTSSFTSLVDNVNRY 16509
Qy 112 ----- 111
Db 16510 DSGKYTLFLENSSGTKSAFVTVRVLDTPSPVNLKVTEITKDSVSIWEPPLDGGSKIK 16569
Qy 112 ----- 111
Db 16570 NYIVEKREATKSYAAVNTCHNSWKIDLOEGCSYFRVTAENEYIGIGLPAQTADPIK 16629
Qy 112 ----- 111
Db 16630 VAEVPPGKITVDVTRNSVLSWTKPEHDGSKIIQIVIVEMQAKHSEKMSACARVKS 16689
Qy 112 ----- 111
Db 16690 QAVITNLQTGEYLFPRVAVNEKGRSDRSLAVPIVAKDLVIEPDVKPAFSSYSVQVQGD 16749
Qy 112 ----- 111
Db 16750 LKIEVPISGRPKTITWTKOGLPLKQTRINVTDSLDTLSIKETHKDDGGQYGITVAN 16809
Qy 112 ----- 111
Db 16810 VVGQKTASIEIVTLDKPDPKGPVKFDDVSAESITLSWNPPLYTGCCOITNIVOKRDTT 16869
Qy 112 ----- 111
Db 16870 TTWPDVWSATVARTTLKVTKLKTGETYQPRIFAENRYGOSFALES DPIVAQYPYKEPDP 16929
Qy 112 ----- 111
Db 16930 GTFPATAISKDSMWIQWHEPVNNGSPVIGYHLERKERNISLWTKVNTKI IHDTOFKAON 16989
Qy 112 ----- 111
Db 16990 LEEGIEYEPVVAENIVGVKASKNSECYVARDPCDPPGTPEPIMVKRNEITLWMTKPY 17049
Qy 112 ----- 111
Db 17050 DGSMTITGYIVEKRDLPDRGMWKASFTNVIETQFTVSGLTEDQRYEFRVIAKNAAGATSK 17109
Qy 112 ----- 111
Db 17110 PSDSTGPIAKDEVELPRISMDPKFRDTIVNAGETFRLEADVHGKPLPTIEWLRGDK 17169
Qy 112 -----PDGMR 116
Db 17170 EESARCEIKNTDFKALLIVKDAIRIDGGQVILRASNVAGSKSPVNVKVLDRPPEGPV 17229

Qy 117 QVRGVT-----QLGG----- 126
Db 17230 QVTGVTSKSLTWSPLQDGGSDISHYVVEKRETSRLAWTVVASEVWVNSLKVTKLLEG 17289
Qy 127 ----- 126
Db 17290 NEYVFRIMAVNKYGVGEPLSAPVLMKNPFVLPGPSKLEVTNIAKDSMTVCNRPDSG 17349
Qy 127 ----- 126
Db 17350 GSEIIGVIVEKRRSGIRWIKCNKRRIIDLRLVTVGLTEDEHEYEFPRVSAENAGVGEPS 17409
Qy 127 -----ACSP----- 130
Db 17410 ATVYKACDPVFKPGPTNAHIVDTTKNSITLAWKPIYDGSSEILGYVVEICKADEEW 17469
Qy 131 ----- 130
Db 17470 QIVTPQTLRVTRFEISKLTEHOEYKIRVCALNKGVLGEATSPGTVPKPEDKLEAPELD 17529
Qy 131 -----TWS----- 133
Db 17530 DSELKGIIVRAGSARIHIPFKGRPMPEITWSREGEFTDKVQIEKGVNYTOLSIDNCD 17589
Qy 134 ----- 133
Db 17590 RNDAGKYLKLENSSGSKSAFVTVKVLDPGPQNLAVKEVRKDSAPLVNPEPIIDGGAK 17649
Qy 134 ----- 133
Db 17650 VKNYVIDKRESTRKAYANVSKSKTSFKVENLTEGAIYFRVMAENEFVGVVPEVTVDA 17709
Qy 134 ----- 133
Db 17710 VKAAEPPSPKVKVTLTDVTSQTSASLWKEPEHDGGSRLGVYVVEMQPQTEKMSIVAESK 17769
Qy 134 ----- 133
Db 17770 VCNVAVTGLSSGQEQYQFRKAYNEKSGSDPRVLGVPVIAKDLTIQPSLKLPFNYSIQAG 17829
Qy 134 ----- 133
Db 17830 EDLKIEIPVIGRPNISWVKDGEPLKOTTRVNVETATSTVLHIKEGNKDDFKYTVTA 17889
Qy 134 ----- 133
Db 17890 TNSAGTATENLSVIVLEKPGPPVGFVDEVSADFVVISWEPAYTGCCQISNIVIVEKRD 17949
Qy 134 ----- 133
Db 17950 TTTTTHMVSATVARTTIKITLKTGETYQPRIFAENRYGKSAPLDSKAVIVQYPFKEPG 18009
Qy 134 ----- 133
Db 18010 PPGTFVTSISKQMLVQMHPEVNDGTTKIIGYHLEKEKNSILWKLNKTPIQDTKPKT 18069
Qy 134 ----- 133
Db 18070 TGLDEGLEYPKVAENIVGIGKPSKVSCECFVARDPCDPPGRPEAIVITRNNVTLKWKPK 18129
Qy 134 ----- 133
Db 18130 AYDGSKITGYIVEKKDLPDGRWMKASFTNVLETEFTVSGLVEDQRYEFRVIARNAAGNF 18189
Qy 134 ----- 133
Db 18190 SEPSDSSGAITARDEIDAPNASLDPKYKDVIVVHAGETVLEADIRGKPIPDVVWVKDGK 18249
Qy 134 ----- 133
Db 18250 ELEETAARMEIKSTIQKTTLVVKDCIRTDGGQYILKLSNVGTSKSIPIITVKVLDPRGSP 18309

QY 134 ----- 133
 Db 18310 GPLKVTGTAECYLAWNPLQDGGANISHYIIIEKRETSRLSWTQVSTEVOALNYKVTKL 18369
 QY 134 ----- 133
 Db 18370 LPGNEYIFRMAVNKYGIGEPLESGPVTACNPKPGPPSTPEVSAITKDSMVVTWARPV 18429
 QY 134 ----- 133
 Db 18430 DDGGTEIEGYILEKRDKEGVRWTKCNKTKLTDLRLRVTLGLTEGHSYEFRAAENAAGVGE 18489
 QY 134 ----- 133
 Db 18490 PSEPSVFRACDALYPPGPSNBKVTDTSRSSVSLAWSKPIYDGGAPVKGYVVEKEAAA 18549
 QY 134 ----- 133
 Db 18550 DEWTTCTPTGLOGKQFTVKLKENTENYFRICAINSEGVGEPATLPGSVVAQERIEPPE 18609
 QY 134 ----- 133
 Db 18610 IELDADLRKVVLRASATLRLFTVIKGRPEPEVKWEKAEGLTDRAQIEVTSFTMLVID 18669
 QY 134 ----- 133
 Db 18670 NVTRFDSGRYNLTLENNSGSKTAFVNVRLDPSAPVNLTIREVKKDSVTLSEPPPLIDG 18729
 QY 134 ----- 133
 Db 18730 GAKITNIVIEKRETRKAYATITNNCTKTTFRIENLOEGSYVFRVLASNEYGIGLPAET 18789
 QY 134 ----- 133
 Db 18790 TEPVKVSEPLPGRVTLVDVTRNTATIKWEKPESDGSKIITGYVVMQTKGSEKSTCT 18849
 QY 134 ----- 133
 Db 18850 QVKLEATISGLTAGBEYVFRVAANKEGRSDRQLGVPVIARDIEIKPSVELPFTFNV 18909
 QY 134 ----- 133
 Db 18910 KAREQLKIDVPFKRQATVNNRKGOTLKETTRVNVSSKTVTSLIKEASKEDVGTYE 18969
 QY 134 ----- 133
 Db 18970 LCVNSAGSITVPITIIIVLDRPGPGPIRIDEVSCDSITISWNPPEYDGGCQISNIVEK 19029
 QY 134 ----- 133
 Db 19030 KETTSTTHIVSQAVARTSIKIVRLTTGSEYQFRVCAENRYGKSSYSESSAVVAEYFPSP 19089
 QY 134 ----- 133
 Db 19090 PGPGTKPVVHATKSTMLVTWQVPVNDGGSRVIGYHLEYKERSILLWSKANKILIADTV 19149
 QY 134 ----- 133
 Db 19150 KVSGLDEGLMYEYRVAENIAGIGKSKCEVPVPCDPGQPEVTNITRKSLSUKWS 19209
 QY 134 ----- 133
 Db 19210 KPHYDGAKITGYIVERRELDPGRWLKCNVTNIOETFEVTELTEDORYEFVRFARNAAD 19269
 QY 134 ----- 133
 Db 19270 SVSEPESTGPIIVKDDVEPRVMDVKFRDVIIVVKAGEVLKINADIAGRPLPVISNAKD 19329
 QY 134 ----- 133
 Db 19330 GIEIEERARTEIISTDNHTLLTVKDCIRRDGTQYVTLKNVAGTRSVAVNCKVLDKPGPP 19389
 QY 134 ----- 133

Db 19390 AGPLEINGLTAECSLSGWRPOEDGGADIDYHHRKKRETSHLAWTICEGELQMTCKVTK 19449
 QY 134 ----- 133
 Db 19450 LLKONEYIFRVTGVNKYGVGEPLESAIKALDPFTVPSPPTSLEITSVTKESMTLCHSRP 19509
 QY 134 ----- 133
 Db 19510 ESDGSEISGYIIERREKNSLRVRVNNKPYVDLVRKSTGLRGECEYEVYVAENAAGLS 19569
 QY 134 ----- 133
 Db 19570 LPSETSPILRAEDPVFLPSPSPKPKIVDSGKTTITIAWVKPLFDGAPITGYTVEYKSD 19629
 QY 134 ----- 133
 Db 19630 DTDWKTISQSLRGTEYTTISGLTTGAEYVFRVKSNNKVCASDPSSDDPOIAKEREPEPLF 19689
 QY 134 ----- 133
 Db 19690 DIDSEMRKTLIVKAGASFTMTVPFRGRPVNVLMKSPDLDLRTAYVDTDSRTSLTIEN 19749
 QY 134 ----- 133
 Db 19750 ANRDSGKYTLTIONVLSAASLTLYVKVLDTPGPPNTITVODVTKESAVLSWDVPENDGG 19809
 QY 134 ----- 133
 Db 19810 APVKYHIEKREASKAWWSVTNNCNRLSYKVTNLOEGAIYYFRVSGENEFQVGIPEATK 19869
 QY 134 ----- 133
 Db 19870 EGVKITEKPSPEKLGVTISKDSVSLTWLKPHEHGGSKIYHYVVEALEKGOKNWKCAV 19929
 QY 134 ----- 133
 Db 19930 AKSTHHVVSGLRENSSEYFRVFAENQAGLSDPRELLPLVLIKEQLEPPEIDMKNPSTHV 19989
 QY 134 ----- 133
 Db 19990 YVRAGSNLKVDPISGKPLPKVTLSDRGVPLKATMRFNTEITAENLTINLKESVTADAGR 20049
 QY 134 ----- 133
 Db 20050 YEITAANSSTTKAFINIVLDRPGPTGPVVISDITEBSVTLKWEPPKYDGSQVTNYI 20109
 QY 134 ----- 133
 Db 20110 LLKRETSTAVMTEVSATVARTMMKVMKLTGTGEYOFRIKAENRFGISDHQACVTVKLP 20169
 QY 134 ----- 133
 Db 20170 YTTGPPSPWTVNTVRESITVGHHEPVSNGSVAVGYHLEMKDRNSILWOKANKLVIRT 20229
 QY 134 ----- 133
 Db 20230 THPKVTISAGLIYEFVRVAENAGVGKPSHPSEPVLAIDACEPPRNVIRITDISKNSVL 20289
 QY 134 ----- 133
 Db 20290 SMOOPAFDGSKITGYIVERRDLDPGRWTKASFTNVTETQFTISGLTONSQYEFVRFARN 20349
 QY 134 ----- 133
 Db 20350 AVGSIENPSEVGPITCIDSYGGPVIDLPLEYEVVKYBAGTSVKLRAGISGKPAPTEW 20409
 QY 134 ----- 133
 Db 20410 YKDDKELOTNALVCVENTTDLASILIKDADRLNSGCEYELKLRNANASATIRVQILDKP 20469
 QY 134 -----CLIT 137
 |||

Db 20470 GPPGPIEFKTVTAETILLWRPPADGGAKITHYIVKEKRETSRVVMSMVSEHLEECIIT 20529
Qy 138 ----- 137
Db 20530 TTKIKGNEYIFRVRANKYGIGEPLESDSVAKNAFVTPGPGIPEVTKITKNSMTVW 20589
Qy 138 ----- 137
Db 20590 SRPIADGSDISGYPLEKXRDKKSLGWFVLKETIRDTROKVTGLTENSQYQYRCVAVNAA 20649
Qy 138 ----- 137
Db 20650 GQGPSESEFYKAADPIDPPGPAKIRIADSTKSSITLGWSKPYDGSVAVTGYVWEIR 20709
Qy 138 ----- 137
Db 20710 QGEBEEMTVSTKGEVRTTEYVWSNLKPGVNYFRVSAVNCAGOGPIEMNEPVOAKDIL 20769
Qy 138 ----- 137
Db 20770 EAPEIDLOVALRTSVIAKAGEDVQVLIIPFKGRPPPTVTRKDEKVLGSDARYSIENTDSS 20829
Qy 138 ----- 137
Db 20830 SLLTIPOVTRNDTGKYLITLTIENGVEPKSSTVSVKVLDTPAACQKQVHVSRGTVLLW 20889
Qy 138 ----- 137
Db 20890 DPPLIDGSPINNVYIEKRDATKRTWSVSHKCSSTFKLIDLSEKTPFFFRVLAENIG 20949
Qy 138 ----- 137
Db 20950 IGECETPEPKAAEVPAPIRDLSMKDSTKTSVILSWTKPDPDGGSVITEYVVERKKGKE 21009
Qy 138 ----- 137
Db 21010 QTWSHAGISKTEIEVSQLEQSVLEFRVFAKNEKGLSDPVTIGTITVKELIITPEVDLS 21069
Qy 138 ----- 137
Db 21070 DIPGAQVTVRIGHNVHLEPYKGPSPISWLKQGLPLKESEFVRFSKTENKITLSIKNA 21129
Qy 138 ----- 137
Db 21130 KKEHGKTVTLDNAVACRIAPITVITLGPSPKPGPIRPFDEIKADSVILSHDVPEDNGG 21189
Qy 138 ----- 137
Db 21190 GEITCYSIEKRETSQTNMKWCSSVARTTFKVPNLVKDAEQFRVAENRYGVSQPLVSS 21249
Qy 138 ----- 137
Db 21250 IIVAKHOFRIPOPPGKPVINYVTSQSLTWDAPVYDGGSEVTGPHVEKKERNILWQV 21309
Qy 138 ----- 137
Db 21310 NTSPISGREYRATGLVEGLDYQFRVYAENSAGLSPPSDPSKFTLAVSPDPPGTPDYIDV 21369
Qy 138 ----- 137
Db 21370 TRETITLKNPPLRDGGSKIYGYSIEKQGNRWRVNCFTDVSECOYTVTGLSPGDRYEF 21429
Qy 138 ----- 137
Db 21430 RIIRARNVGTISPPSQSGIIMTRDENVPPIVEFGPEYFDGLIISKGESLRKALVQGRP 21489
Qy 138 ----- 137
Db 21490 VPRVTWFKDGEIEKRMMEITNVLGSTSLFVRDATRDHRGVYTVBEAKNASAKAEIKV 21549
Qy 138 ----- 137
Db 21550 KYQDTPGKVVGPIRFTNITGEKMTLWMDAPLNDGCAPIHYIIIEKRETSRLAWALIEDKC 21609

Qy 138 ----- 137
Db 21610 EAQSYTAIKLINGEYQFRVSAVNKEGVRPLSDPVAQIQYTVDPAGIPEPSNITGN 21669
Qy 138 ----- 137
Db 21670 SITLTWARPESDGGSEIQYILERREKSTRVWKVISKRPISETRFKVTGLTEGNEYEFH 21729
Qy 138 ----- 137
Db 21730 VMAENAAVGVPASGISRLIKREBPVNPBPPTVVKVTDTSKTTVSLSEWSKPVFDGGMELI 21789
Qy 138 ----- 137
Db 21790 GYIEMCKTDLGDWHKNAEACVKTTRYVTDLQAGEEYKFRVSAINGAGKGDSEVTGTI 21849
Qy 138 ----- 137
Db 21850 KAVDRLTAPELIDANFKQTHVVRAGASIRLFIAYQGRPTPTAVWSKPDNSLSRADIHT 21909
Qy 138 ----- 137
Db 21910 TDSFSTLTVENCMNRNDAGKTYLTVENNNGSKSITFTVKVLDTPGPPGPIITFKOVTRGSAT 21969
Qy 138 ----- 137
Db 21970 LMWDAPLLDGGARIHHYVVEKREASRSQVISEKCTRIQIFKVNDLAEGVPPYFRVSAVN 22029
Qy 138 ----- 137
Db 22030 EYGVGPEYEMPEPIVATEQPAPRRRLDVTDSKSSAVLAWLKPDDHGGSRITGYLLEMRO 22089
Qy 138 ----- 137
Db 22090 KGSDLWVEAGHTKOLTFTVERLVEKTEYEPFRVAKNDAGYSEPREAFSSVIIKEPQIEPT 22149
Qy 138 ----- 137
Db 22150 ADLTGINTQLITCKAGSPFTIDVPIISCRPAPKVTWLEEMRLKETDRVSITTTKDRTTLT 22209
Qy 138 ----- 137
Db 22210 VKDSMRDGSRYFLTLENTAGVKFTSVTVVVGIPGPTGPIEVSVSAESCVLWSGEPK 22269
Qy 138 ----- 137
Db 22270 DGGTEITNIVIEKRESGTTAMQLVNSSVKRTQIKVTHLTMYMEYSFRVSENRFVSKP 22329
Qy 138 ----- 137
Db 22330 LESAPIIAEHFPVPPSPAPTREPVYHVSAVANSIRWEERPYHDGSKIIGYMWVEKKERNITL 22389
Qy 138 ----- 137
Db 22390 WVKENKVCLECNKYVTGLVEGLEYPQRTVALNAAVSKASEASRPIMAOQNPVDAPGRPE 22449
Qy 138 ----- 137
Db 22450 VTDVTRSTVSLIMSAPAYDGGSKVGVYIIERKPVSEVGDGRWLKCNVTIIVSDNFTVTAL 22509
Qy 138 ----- 137
Db 22510 SEGDTYFRVLAKNAAGVISKGSSESTGPTCRDEYAPPKAELDARLHGLDVTIRAGSDLV 22569
Qy 138 ----- 137
Db 22570 LDAAVGKPEPKIITWKDKELDLCEKVSLOYTQGRATAVIKFCDRSDSGKYTLTVKNAS 22629
Qy 138 ----- 137
Db 22630 GTKAVSVNVKVLDSPPGCGKLTVSRVTOEKCTLAWSLPQEDGGAEITHYIVERRETSRLN 22689

Qy	138	-----	137
Db	22690	WVIVEGECPTLSVVVTRLKNNKEYIFRVRANKYGPVPESEPIVARNSFTIPSPGP	22749
Qy	138	-----	137
Db	22750	EEVGTGKEHIIQWTKPESDGGNEISNLYLVDKREKSLRWRVNDKYVVYDTRLKVTSLM	22809
Qy	138	-----	137
Db	22810	EGCDYQFRVTAVNACGNSEPSERSNFISCREPSYTPGPPSAPRVVDTTKHSISLAWTKPM	22869
Qy	138	-----	137
Db	22870	YDGGTOIVGVLEMOEKDQWYVHTNATIRNTEFTVPLKMGQKYSFRVAANVKGMS	22929
Qy	138	-----	137
Db	22930	EYSIESIAEIPVERIEIPDLELADDLKKTVIRAGASLRMLVSVSRPPPVITWSKQIGD	22989
Qy	138	-----	137
Db	22990	LASRAIIDTTESYLLIVDKVNRVDAGKYTIEAENOSGKSATVLVKVYDTPGPCPSVKV	23049
Qy	138	-----	137
Db	23050	KEVSRDSVTITWEIPTIDGAPINNVIVEKREAAAFKTVTTKCSKTLYRISGLVEGTM	23109
Qy	138	-----	137
Db	23110	HYPRVLPENIYGIGPCETSDAVLVSEVPLVPAKLEVVDTKSTVTLAWEKPLYDGSRL	23169
Qy	138	-----	137
Db	23170	TGYVLEACKAGTERMMKVWLKPTVLEHTVTSLEGEQYLFIRIAQNEKGVSEPRETVTA	23229
Qy	138	-----	137
Db	23230	VTQDLRLPTLDLSTWPKQTHVPAGRPVELVPIAGRPPPAASWFFAGSKLRESERV	23289
Qy	138	-----EDTG-----	141
Db	23290	VETHKVAKLTIRETTTIRDGEYTLKQVTTSETIKVILDKPGPTGPIKIDEIDA	23349
Qy	142	-----	141
Db	23350	TSITISWEPPELDGGAPLSGYVWEQDAHRPGWLPVSESVTRSTFKFTRLTEGNEYVFRV	23409
Qy	142	-----	141
Db	23410	AATNRFIGSYLQSEVIECRSSIRIPGPPETLQIFDVSRDGMTLTWYPPEDDGGSQVTGY	23469
Qy	142	-----	141
Db	23470	IVERKEVRADRWRVNVKVPVMTVRYSRSTGLEGEYEHVRVTAINARGSGKPSRSPKIVA	23529
Qy	142	-----	141
Db	23530	MDPIAPGKPNQPRVTDTRTSVSLAWSVPEDEGSKVTGYLIEQKVDOHEWTKCNTTP	23589
Qy	142	-----	141
Db	23590	TKIREYTLTHLPQGAEYFRVLACNAGGPEAEVPGTVKVKTEMLEYPDYELDERYQEGI	23649
Qy	142	-----	141
Db	23650	FVROGGVIRLTIPIKGPFPICKMTKEGODISKRAMIATSETHTELVIKEADRGDSGYD	23709
Qy	142	-----	141
Db	23710	LVLENKCGKAVYIKVRVIGSPNPEGPLEYDDIQVRVRSWRPPADGGADILGYILE	23769
Qy	142	-----	141

Db	23770	RREVPKAAWVTIDSRVRGTSLSLVKGLKENVEYHFRVSAENQFGISKPLKSEEPVTPKTPL	23829
Qy	142	-----	141
Db	23830	NPPEPPSNPPEVLVDTWKSSVLSWSRPRKDDGSGRVTTYIERKETSTDKWRHKNKTQITT	23889
Qy	142	-----	141
Db	23890	TMYTUTGLVPDAEYQFRIIAQNDVGLSETSPASEPVPVKDPDFKPSQPOGELEILSISKDS	23949
Qy	142	-----	141
Db	23950	VTLOWKEPCDGGKEILCYWVEYRQSGDSAMKSNKERIKDKQFTIGGLLEATEYEFVRV	24009
Qy	142	-----	141
Db	24010	AENETGLSRPRTAMSIKTKLTSGEAPGIRKEMKDVTTKGEAAQOLSQOIVGRPLPDIKM	24069
Qy	142	-----	141
Db	24070	YRFGKELIOSRKYKMSDGRTHLTVMTEEOEDEGVYTCIATNEVGEVETSSKLLQATP	24129
Qy	142	-----	141
Db	24130	QPHPGYPLKEYGAVGSTLRHLHVMIYIGRPVPMTWPHGOKLLQNSENTIENTEHTHL	24189
Qy	142	-----	141
Db	24190	VKNVORKTHAGKYKVQLSNVFGTVDAILDVEIQDKPDKPTGPIVIEALLKNSAVISMKP	24249
Qy	142	-----	141
Db	24250	PADGGSMITNVYVEKCEAGAEQWLVSSAISVTTCRIVNLTENAGYVFRVSAQNTFGI	24309
Qy	142	-----	141
Db	24310	SDPLEVSVVIIKSPFEPGAPKPTITAVTKDSCVWAKPPASDGGAKIRNYYLEKREK	24369
Qy	142	-----	141
Db	24370	KONKWSVTTEIRETVSVKQNLIEGLEYPVAVKCNLGGSESESEISEPITPKSDVPDQ	24429
Qy	142	-----	141
Db	24430	APHFKEELRNLNVRYQSNATLVCKVTGHPKPIVKWYRQGEIIADGLKYRIOEFKGYHQ	24489
Qy	142	-----	141
Db	24490	LIIASVTDDATVYQVRATNOGGSVSGTASLEVEVPKIHLPKLTLEGMAHALRGEVVS	24549
Qy	142	-----	141
Db	24550	IKIPFSGKDDPVITWQKGODLIDNNCHYQVIVTRSFVSLVFPNGVERKDAGFYVVCANR	24609
Qy	142	-----	141
Db	24610	FGIDQKTVELDVADVPDPPRGVKVSDASRDSVNLWTWTEPASDGGSKITNYIVEKCATTA	24669
Qy	142	-----	141
Db	24670	RWLRVGOARETRYTVINLFGKTSYQPRVIAENKFGLSKSPSESEPTITKEDKTRAMNYDE	24729
Qy	142	-----	141
Db	24730	EVDETREVSMTKASHSTKELYEKYMIADPLGRGEFGIVHRCVETSSKTYMAKFKVKVG	24789
Qy	142	-----	141
Db	24790	TDQVLVKKEISILNIARHNLHSHSFESMBELVMIFISGLDIFERINTSAFELNER	24849
Qy	142	-----FDL-----	144
		:	

Db 24850 EIVSYHVCEALQFLHSHNIGHDIRPENIIYQTRRSSTIKIIEFQARQLKPGDNFRL 24909
 QY 145 ----- 144

Db 24910 LFTAPEYAPEVHQHDVYVSTATDMWSLGLTVYVLLSGINPFLAETNOQIIENIMNAEYTF 24969
 QY 145 ----- 144

Db 24970 DEAPKEISIEAMDVDRLLVKERKSRMTASEALQHPMLKQKIERVSTKVIRTLKHRRY 25029
 QY 145 ----- 144

Db 25030 HTLIKDLNMVSAARISCGAIRSQKGSVAKVKVASIEIGPVSGQIMHAVGEGGHVK 25089
 QY 145 ----- 144

Db 25090 YVCKIENYDOSTQVWTYFVGRQLENSKEYEITYEDGVAILYVVKDITKLDGTYRCKVND 25149
 QY 145 ----- 144

Db 25150 YGEDSSVAELFVKGVREVDYCRRTMKIKRRTDTWRLLEPPEFTPLYNKTYVGEN 25209
 QY 145 ---GVTI 148
 |||||
 Db 25210 VRFGVTI 25216

RESULT 11
 AAR44929
 ID AAR44929 standard; Protein; 15281 AA.
 XX
 AC AAR44929;
 XX
 DT 08-JUL-1994 (first entry)
 XX
 DE T. niveum Cyclosporin synthetase.
 XX
 KW Enzyme; cyclosporin; synthetase-like activity; Tolypocladium niveum;
 KW T. inflatum GMS; biosynthesis; vector; cyclosporin synthetase.
 XX
 OS Tolypocladium niveum.
 XX
 PN EP578616-A.
 XX
 PD 12-JAN-1994.
 XX
 PF 05-JUL-1993; 93EP-0810474.
 XX
 PR 09-JUL-1992; 92AT-0001403.
 PR 08-MAR-1993; 93AT-0000437.
 PR 29-APR-1993; 93CH-0001310.
 PR 04-MAY-1993; 93CH-0001375.
 XX
 PA (SANO) SANDOZ LTD.
 PA (SANO) SANDOZ PATENT GMBH.
 PA (SANO) SANDOZ-ERFINDUNGEN VERW GES MBH.
 PI Leitner E, Schneider E, Schoergendorfer K, Weber G;
 XX
 XX WPI; 1994-010432/02.
 DR N-PSDB; AAQ54386.
 XX
 PT Isolated DNA sequence - which codes for enzyme having cyclosporin
 PT synthetase like activity
 XX
 XX Claim 1; Page 41-84; 93pp; English.
 XX
 CC This sequence represents an enzyme which has cyclosporin synthetase-
 CC like activity. This sequence was isolated from Tolypocladium niveum
 CC (formerly known as T. inflatum GMS). This enzyme catalyses the
 CC peptide biosynthesis of cyclosporins and structurally related
 CC molecules. This sequence may be used for the production of
 CC cyclosporin by transforming a vector containing this sequence in

CC to a recombinant host. This allows effective production of anti-
 CC biotic cyclosporin or its derivatives.
 XX
 SQ Sequence 15281 AA;
 Query Match 34.2%; Score 264; DB 15; Length 15281;
 Best Local Similarity 0.8%; Pred. No. 0.3;
 Matches 102; Conservative 24; Mismatches 21; Indels 12475; Gaps 32;
 QY 1 AAGGIL----- 6
 |||||
 Db 2098 AACHVLEIGTGTGMVLFNLQAGLKSYYIGLEPSQSAVFVNKAAQTFPGLEGKAOVHVGT 2157
 QY 7 ----- 6
 Db 2158 AMDTGRLSALSPDLIVNSVAQVFPSPREYLAEVVEALVRIPGVRRIFFGDMRTYATHKDF 2217
 QY 7 -----HLELL----- 11
 |||||
 Db 2218 LVARAVHTNGSKVTRSKVOQEVARLEELLDVDPAFFTSLSKESIEIEHVEILPKNM 2277
 QY 12 ----- 11
 Db 2278 KVNELSSRYGAVLHNRHNNQNSRIHKINAESWIDFASSQMDRQGLARLLKENKDAE 2337
 QY 12 ----- 11
 Db 2338 STAVFNIPYSKTIIVERHIAKSLADDHGDGDDTHSSIDGVAWISAAREKASQCPSLDVHDLV 2397
 QY 12 ----- 11
 Db 2398 OLAEDAGRFEVSWARQSONGALDVFPHFQPTENESRALVDFTDYKQQARSUTNRP 2457
 QY 12 ----- 11
 Db 2458 LORVESRRIEAQVREQLQVLLPAYMIPARIVVLQNMPLNTSGKVDKELTLRAKVTAAAT 2517
 QY 12 ----- 11
 Db 2518 PSSELVAPRDSIEAIIKEFKDVLGVEGIDTNFNVGHSLLATKLAARLSQLNAQIA 2577
 QY 12 ----- 11
 Db 2578 VKDIFDRPVIADLAATIOQDTTEHNPIPTSYTGPVEQSAQGRWLFLDQLNVGATWYLM 2637
 QY 12 ----- 11
 Db 2638 PFAVLRGLVWSALAAALLALEERHETLRTTFIEQEGIMQVHPFAPKELRVIDVSGE 2697
 QY 12 ----- 11
 Db 2698 EESTIQILEKEQTTPFNLAEPGFRLLKTGDESHILSTMHHAISDGWSVDIFQOEI 2757
 QY 12 ----- 11
 Db 2758 GQFYSAILRGHDPLAQIAPLSIQYRDFATWQROIQVFAEHRRLAYWTKOLANKPAELL 2817
 QY 12 ----- 11
 Db 2818 TDFKRPPLMSGRAGEIPVVVDGLIYEKLQDFCIRQVTAFTVLLAAPRAAHYMTGTEDA 2877
 QY 12 ----- 11
 Db 2878 TICTPIANRRPELEGLIGFFVNTQCMRITVDVEDSFETLVHQVRETTLAAHANQDVPE 2937
 QY 12 ----- 11
 Db 2938 QIVSNILPGSSDTSRNPVLQMFALHSQONLKGVRLEGIEEIIISIAETTRFDIEFHYQ 2997
 QY 12 ----- 11
 Db 2998 EAERLNGSIVYAADLFPETIQSVITIFQILOGLGEPMVPVSMALDGLSRLSTGL 3057

Qy 12 -----VAGPDV----- 18
Db 3058 LHPQOTDPCDASVQIFKQOAVNPDVIAVRDESTRLSYADLRKSDQVACWLSRRGIA 3117
Qy 19 ----- 18
Db 3118 PETFAILAPRSCETIVAILGLVKANLAYPLDVNVPASRLAAILSEVSGSMLVLVGAET 3177
Qy 19 ----- 18
Db 3178 PIPEGMAEATIRITEILADAKTDDINGLAASOPTAASLAYVIFTSGSTRPKGVMEHR 3237
Qy 19 ----- 18
Db 3238 GIVRLTKOTNITSKLPESFMAHJSNLAFDASVWEVFTLLNGGLVLCIDYFTLLESTAL 3297
Qy 19 ----- 18
Db 3298 EKVFQORVNVALLPPALLKQCLDNLSPALVKLTLSVLYIGDRLDASDAAKARGLVQTQAF 3357
Qy 19 ----- 18
Db 3358 NAYGPTVMSTIYPIAEDPPINGVPIGHAVSNSGAFVMDQNOQITPPGAMGELIVTGD 3417
Qy 19 ----- 18
Db 3418 GLARGYTTSSLNTRGFINDIDGEOVRAYRTGDRVRYRPKDLOIEFFGRIDHQVKIRGHR 3477
Qy 19 ----- 18
Db 3478 IEPAEVEYALLSHDLVDAAVTHSOENQDLEMVGFVAARVADVREDESSNQVQEWQTHF 3537
Qy 19 ----- 18
Db 3538 DSITAYADITTIQQSLGRDFMSWTSMDGSLIKKSQWELDDTMRSLDSDPPGHVLEV 3597
Qy 19 ----- 18
Db 3598 GTGTGMVFNLRGGLQSYVGLSPSATAFVNKAAKSPFPLEDRIRVEVGTATDIDL 3657
Qy 19 ----- 18
Db 3658 GDDLHAGLVVNSVAQYFFSQDYLAQLVRDLTKVPGVERIFFGDMRSHAINRDFLVARAV 3717
Qy 19 ----- 18
Db 3718 HALGDKATKAEIQREVVRMESEDELLVDPAPFTSLTQVENIKHVEILLPKRMRATNELS 3777
Qy 19 ----- 18
Db 3778 SYRYAAVLHVNDLAKPAHKVSPGAWVDFAAATKMDRDALIRLLRGTIKISDHIAIANIPNSK 3837
Qy 19 ----- 18
Db 3838 TIVERTICESVYDLGGDAKSDNRVSWLSAARNAVKVASLSAIDLVDIAQEAAGRVEIS 3897
Qy 19 ----- 18
Db 3898 CAROWSQNGALDAVFHHLGPSPOSSHVIDFLTDHQRPPEALTNHPLHRAQRRRVERQI 3957
Qy 19 ----- 18
Db 3958 RERLOTLLPAYMPIAQIMVLDKLPNANGKVDKRQLTORAQTVPKAKQVSAPVAPRTEIE 4017
Qy 19 ----- 18
Db 4018 RVLQBFSDVLGVDIGIMENFFDLGGHSLMATKLAARISRRLETHSVKEIFDHPRVCDL 4077
Qy 19 ----- 18
Db 4078 VLIVOGSAPHDPVSTKYTGVPQSFQAGRLWFLDQNFQATWYLMPLAVRLRGAMNVH 4137
Qy 19 ----- 18

Db 4138 ALTAALLALERRHELLRTTTFYEQNGVMQKVPVWTETLRIIDLNSGDGDYLYTLKKEOT 4197
Qy 19 ----- 18
Db 4198 APFHLETEPGWRVALLRLGPDYILSVMMHHIISDGWSVDVLFOELQOFYSTAVKGHDPL 4257
Qy 19 ----- 18
Db 4258 SOTTPLPIHYRDPALWOKKPTQSEHERQLOYWVEQLVDSAPAEALLTDLPRPSILSGOAG 4317
Qy 19 ----- 18
Db 4318 EMSVTIEGALYKNLEEFRCVRHRTVSFVLLAALRAAHVRLTGSSEDATIGTBIANRNPRL 4377
Qy 19 ----- 18
Db 4378 EQIIIGFFVNTOCIRITVNEDETFESLVQOVRSTATAFAHODVPFEKIVSTLLPGSRDAS 4437
Qy 19 ----- 18
Db 4438 RNPLVQLMFAVHSOKNLGELKENAHSEVVPTEITTRFDLEFHLFOQDDKLEGSILYSTD 4497
Qy 19 ----- 18
Db 4498 LFEAVSVQSLLSVFQEILRRGLNGPDVPISTLPLQDGIQDLQROGLLDVOKTEYPRDSSV 4557
Qy 19 ----- 18
Db 4558 VDVFEQVINSIPDSIALIHGSEKLSYAQLDRESORVARWLHRHSFSSDTLIAVLAPRCE 4617
Qy 19 ----- 18
Db 4618 TIIAPLGILKANLAYPLDVKAPAARIDAIYSSLPGNKLILLGANVTPPKLOEAAIDFVP 4677
Qy 19 ----- 18
Db 4678 IRDTFTLLTDLQDPTIERPSAQSLAYAMFTSGTGRPKGMVQHRNIVRLVKNSNV 4737
Qy 19 ----- 23
Db 4738 AKQPAARIAHISNLAFLDASSWEIYAPLLNGGAVICADYFTTIDPQALOETFOEHEIRGA 4797
Qy 24 ----- 23
Db 4798 MLPPSLLKQCLVQAPDMISRDLILPAAGDRFSSVDALQARLVGSGVFNAYGPTENTILS 4857
Qy 24 ----- 23
Db 4858 TIYNVAENDSFVNGVPIGSAVNSGAYIMDKNOQLVPAGVMGELVVTGDCARGYMDPKL 4917
Qy 24 ----- 23
Db 4918 DADRFIQTVNGSEQVRAIRTGDRVRYRPKOFQIEFFGRMDQOIKIRGRIEPAEVEQNF 4977
Qy 24 ----- 23
Db 4978 LNDGFVEDVAIVIRTPENQEPWVAFVTKGDNARSAREEATTQIEGWEAHEGGAYANIE 5037
Qy 24 ----- 26
Db 5038 EIESEALGYDFMGWTSMYDGTIDEIDKEMREMLNDTMRSLLDGKPGARVLEVGTTGTMIMF 5097
Qy 27 ----- 30
Db 5098 NLGRSQGLERYIGLEPAPSAAEFVNNAKSPFLAGRAEVHVGTAADVGTLOGLTSDMAV 5157
Qy 31 ----- 30
Db 5158 INSVAQYFPTPEYLAETIKSLVQVPMKRIYLGDMRSMWNRDPAAARAAYSADNASKO 5217
Qy 31 ----- 30

Db 5218 RVQKMMLEEBEELLVDPAFFALTASOLOQRIQHVEILLPKRMKATNELSSRYAAVLH 5277
Qy 31 ----- 30
Db 5278 ISDEPLPIYKIDPEAMINFGSRLTRELAQVLKENENAESVAISNIPYKTVVERHIVR 5337
Qy 31 ----- 30
Db 5338 SLOEDANAPEESMDGSDWISAVRTAQOCHTILSASDLFDIAEDAGFRVEVSWARQHSQ 5397
Qy 31 -----LTV----- 33
|||
Db 5398 GALDAVPHLKPATEDSRVLKIFPTDHQGRPLKSLTNQPLLPQAQSRRAELLIREGLQTL 5457
Qy 34 ----- 33
Db 5458 PPMWPSQITLIDRMPLNANGKVDRLARRAKITQKSKPVEDIVPPRNSVEATVCKGFT 5517
Qy 34 ----- 33
Db 5518 DVLGVEGITDNFFNLGGHSLMATKLAARLGRQLNTRISVRDVFDPQPVADLAAVIQRS 5577
Qy 34 -----LNIG----- 38
|||
Db 5578 APHEPIKPADYTGVPVQSPAQGRFLWFLDQNVGATWYMLPLGIRLHGLSRVDALATAISA 5637
Qy 39 ----- 38
Db 5638 LEORHEPLRTTFHEEDGVGVQVODHRPKDLRIIDLSQPKDAYLAVLKHQETTLFDLAT 5697
Qy 39 ----- 38
Db 5698 EPGWRVALIRLGBEEHILSIVMHHIISDGHVSEVLPDEMHRFYSSALRQDPMEQILPLP 5757
Qy 39 ----- 38
Db 5758 IQYRDFAAQKTEBQAEHORQLDYMTHELADSTPAELLTDLPRPSILSGRANELPLTIE 5817
Qy 39 ----- 38
Db 5818 GRHLKLRACFRVHOATPFVILLAAARAHYRLTGAEDATLGTPIANRNRELEENMIGFF 5877
Qy 39 ----- 38
Db 5878 VNTQCMRIAENDNFESLVRVRSTATSAFANQDVPFESIVSSLLPGSRDASRNPLVQV 5937
Qy 39 ----- 38
Db 5938 ILAVHSQDLGLKLTLEGLRDEAVDSAITRFDVEFHLFEHADRLSGSVLYAKELFKLRTI 5997
Qy 39 ----- 38
Db 5998 ESVSVFLETLRALDQPLTFLAVLPLTDGVGGEIASKGLLDVPRDTPYPRDANIVEVQFH 6057
Qy 39 ----- 38
Db 6058 VRATPDAIVKDATSILTYAQDQSDRLALWLSRRHMPETLVGLAPRSCETIIAMFG 6117
Qy 39 ----- 38
Db 6118 IMKANLAYPLDINSAPARLSILSAVDGNKLVLLGSGVTAQEQENPEVAVGIEILAG 6177
Qy 39 ----- 38
Db 6178 TGLDKTQGSNARPSATSLAYVIFTSGTGPKGMVMEHRSVTRLAKPSNVISKLPQGARV 6237
Qy 39 ----- 38
Db 6238 AHLANIAPDASIWEIATTLNGATLVCLDYHTVLDCTRLKEVFERESITVVTMLPALLKQ 6297
Qy 39 ----- 38
Db 6298 CVAEIPETLAHLDLLYTCGDRVGGHDAMRARSVLKIGMFGYGPPTNTVISTIVEVDADE 6357

Qy 39 ----- 38
Db 6358 MFVNGVPIGKTVNSGAYVMDRNOQLVPBGVVGELVVTGDGLARGYTDPSLKNRFIYIT 6417
Qy 39 ----- 38
Db 6418 VNGESIRAYRTGDRVRYRPHDLQIEFFGRMDQOVKIRGHRIEPGEVESALLSHNSVQDAA 6477
Qy 39 ----- 38
Db 6478 VVICAPADODSAGEMVAFVAARNTEDEDTOBEEAVDOVOGWETHFETAAYSEVKDIROSE 6537
Qy 39 ----- 38
Db 6538 VGNDFMGWTSMDGSEIDKIDMHEWLNDTMRMILDAREPHVLEIGTGTGMVMFNLAACP 6597
Qy 39 ----- 38
Db 6598 GLQGYVGFEPSPKSAQFVNDAASPALKDGRSIVHVGTATDINKAGPIQPRLVVINSVA 6657
Qy 39 ----- 38
Db 6658 QYFPTPEYLFRRVVEALVQIPSVVERIVFGDMRTNAINRDFVASRALHTLQOKANKRLVRQM 6717
Qy 39 -----ELLRDP----- 45
|||
Db 6718 IYELEANEHEELTDPAFFTSLRGLGEKIKHVEILLPKTKMATNELSKYRYAAVLHVGRSR 6777
Qy 46 ----- 45
Db 6778 EQSTIHQVSPNAWIDFAADGLDROTLINLLKEHKDAGTVAIGNIPYSKYTIVERFVNKSLS 6837
Qy 46 ----- 45
Db 6838 EDDMEGQNSLDCSAAWAAVRMAAOSCPSLDAMDVKEIAQEAQYQVEVSWARQWSONGAL 6897
Qy 46 -----LGAQ----- 49
|||
Db 6898 DAIFHHFEPKEGARTLIEFPTDYEGRNVTLTNRPLNSIQSRRLGTQIREKLQTLPLPV 6957
Qy 50 ----- 49
Db 6958 MIPSRIWLDQMPVNNNGKIDRKELVRRAIIVAPKPRSAATRVAPRNEIBAILRDEFEDVL 7017
Qy 50 ----- 49
Db 7018 GTEVSVLDNFFDLGGHSLMATKLAARVSRRLDAHISIKDVFQOPVLADLAASIQRESAPH 7077
Qy 50 ----- 49
Db 7078 EPIQRPYTGPAEQSFAQGLWFLDQNLGATWYLMPLAIRIGQURVAALSAAFLALER 7137
Qy 50 ----- 49
Db 7138 RHETLRTTPEESDGVGVQIVGEARNSDLRVHDVSTGDDGEYLEVLRREQTVPDLSSEPG 7197
Qy 50 FRVHLVK----- 56
: : : : :
Db 7198 WRVCLVKTGEEDHVLIVMHHIYDGSVDILRGELGQFYSAALRGQDPLLHANPLPIQY 7257
Qy 57 ----- 56
Db 7258 RDEAANQREAKQVEEHORQOLGYWSKQLVDSTPAELLTDLPRPSILSGRAGSVDTIEGSV 7317
Qy 57 -----MWILT----- 61
: : : : :
Db 7318 YGALQSFCTRSTRVTTEFWLLTVFRIAHFRLTAVDDATIGTPIANRNRELETLVGCFWNT 7377
Qy 62 ----- 61
Db 7378 QCMRISIADDDNFEGLRVQRNVATAAYANQOVPPERIVSALVPGSRNTRNPLVQLMFA 7437

QY 62 ----- 61
Db 7438 VQSVEDYQVRLGLESVMFGEASTRFDMFHLVPGQKLTGSLVYSSDLFEOGTQNF 7497
QY 62 ----- 61
Db 7498 VDIQOECLRSVLDQPLTIPSVLPFSNAISNLESLLLEMTSDYPRDRTVVDLFRQAAI 7557
QY 62 ----- 61
Db 7558 CPDSIAVKDSSQLTYAQLDEOSDRVAAWLHERHMPAESLVGLSPRSCETIIAYFGIMK 7617
QY 62 ----- 61
Db 7618 ANLAYPLDVYAPDARLAAILDVGERLLLLGAGVPOPGIQI PRLSTAYIAEALSHATT 7677
QY 62 ----- 65
Db 7678 VDVTSI PQPSATSLAYVIFTSGTGKPGVMIEHRGIVRLVRDTNVNVPESGSALPVSH 7737
QY 66 ----- 65
Db 7738 FSNLAWDAATWEIYTVLNGGTVCIDRDTMLDIAALNSTPERKENVRAAFPTPAFLKQCL 7797
QY 66 --APNITANL----- 73
Db 7798 AETPELVANLEILHTAGDRLDPGDANLAGKTAKGGIFNVLGHNTENTAYSTFYPPVGEETF 7857
QY 74 ----- 73
Db 7858 VNGVPGRGISNSHAYIIDRHOKLPAGVMGELLTGDGVARGYDTSALNKDRFVYIDIN 7917
QY 74 ----- 73
Db 7918 GKSTWSYRTGKARYPRDQGLEFFGRMDQMKIRGVRIEPEGEVELTLLDHKSVLAAVTV 7977
QY 74 ----- 73
Db 7978 VRRPPNGDEPIAFITIDAEDDVQTHKAIYKHLQILPAYMIPSHLVILDQMPVTDNGKV 8037
QY 74 ----- 73
Db 8038 DRKDALARAOQVQKRSTAAVPRDEVEAVLCEEYSNLEVEVGITDGFGLGGHSLLA 8097
QY 74 ----- 73
Db 8098 TKLAARLSQLNTRVSKOVFDQPI LADLADIIRGSHRHPPIPATPYTGPVEQSFQAQR 8157
QY 74 ----- 73
Db 8158 LWFLEQLNGASWYLMPPAIRMRGPLOTKALAVALNALVHRHEALRTTFEDHDGVGVQVI 8217
QY 74 ----- 73
Db 8218 QPKSSODLRIIDLSDAVDDTAVLAALKREOTTAFDLTSEPGWRVSLLLRLGDDDDYILSIWM 8277
QY 74 ----- 73
Db 8278 HHIISDGWTVDVLRQELGOFYSAAIRGOEPLSOAKSLPIQYRDFAVWQEQNOIKEQAKQ 8337
QY 74 ----- 73
Db 8338 LKYWSQOLADSTPCFELTDLPRPSILSGEADAVPMVIDGTIVYQLLTDTCRTHQVTSFVSL 8397
QY 74 ----- 73
Db 8398 LAAFRTAHYRLTCTLDATVGTPIANRNRPELEGLIGFFVNTQCMRAISETETFESLVQQ 8457
QY 74 ----- 73
Db 8458 VRLTTTEAFANODVPFEQIVSTLLPGSRDTSRNPVQVMFALOSQODLGRIOLEGMTDEA 8517
QY 74 ----- 73

Db 8518 LETPLSTRLDLEVLHFOEVLKLSGLLYSTDLFEVETIRGIVDVFLFIILRRGLEOPKQRL 8577
QY 74 ----- 73
Db 8578 MAMPIITDGIITKLRDOGLLTVAKPAYPRESVIDLFRQOVAAAPDAIYWDSSSTLYADL 8637
QY 74 ----- 73
Db 8638 DGOSNKLAWLQORNMAPETLVAVFAPRSCLTIVAFGLVKANLAYLPLDVNAPARIEA 8697
QY 74 ----- 73
Db 8698 ILSAVPGHKLVLQAHGPELGLTMADETQIDEALASSSSGDHEQTHASGPTATSLAYV 8757
QY 74 ----- 73
Db 8758 MFTSGSGTGKPGVMIDHRSIIRLVQNSDVVATLPTPVRMANVSNLAFDISVQEIYTTALL 8817
QY 74 ----- 73
Db 8818 GGTLVCLDYLTLTLLDSKILYNVFEAQVNAAMFTPVLLKQCLGNMPAIISRLSVLFNVGDR 8877
QY 74 ----- 73
Db 8878 LDAHDAVASGLIQDAVYNAYGPTENGMSQSTMYKVDNPEFVNGVPIGRSITNSGAYVMD 8937
QY 74 ----- 73
Db 8938 GNQOLVSPGVNGEIVTGDGLARGYDSDALDEDRFVHVTTIDGSENIKAYRTGDRVYRPX 8997
QY 74 ----- 73
Db 8998 DFEIEFFGRMDQVKIRGHRIEPAEVEHALLGHDLVHDAVAVLRKANOBPEMIAFITSQ 9057
QY 74 ----- 73
Db 9058 EDETIEQHSNKVQGVGHEHFDVSRYADIKDLDTSTFGHDFLGWTSWYDVIDPVNEMKE 9117
QY 74 ----- 78
Db 9118 WLDETTASLLDNRRPPGHILEIGAGTGMLSNLKGVDGLQKVGLDPAAPSAIIFVNEAVKS 9177
QY 79 ----- 78
Db 9178 LPSLAGKARVLVGTALDIGSLDKNEIQPELVWINSVAQYFPTSEYLIKVKVAVVEVPSVK 9237
QY 79 ----- 78
Db 9238 RVFFCDIRSOALNRDFLAARAVRALGDNASKEQIREKIAELEBSEBELLVQPAFFVSLRS 9297
QY 79 ----- 78
Db 9298 QLPNIKHVEVLPLKMKATNELSSYRYAAVLHISHNEEQLLIQDIDPTAWVDFAATOKDS 9357
QY 79 -----SVC 81
Db 9358 QGLRNLQOGRDDVMIAVGNIPYSKTIIVERHIMNSLDQDHVNSLDGTSWISDARSAAC 9417
QY 82 ----- 81
Db 9418 TSFDAPALTQAKBEGFRVELSWARQSONGALDAVPHRLATDANCERSRVLVHPTDQ 9477
QY 82 ----- 81
Db 9478 GRQLRTLNRPLQRAQSRRIESQVFEALQTPALPAYMIPSRILVLPQMPNTANGKVDKQOL 9537
QY 82 ----- 81
Db 9538 ARRAQVAVKAVSARVAPRNDTEIVLCEEYADILGTEVITDNFDMGHSMLATKAA 9597
QY 82 ----- 81

Db 9598 RLSRLDTRVTVKEVDFKPVLDLAASIEOGSTPHLPPIASSVYSGPVEQSYAQGLWFLD 9657
Qy 82 ----- 81
Db 9658 QFNLNATWYHMSLAMRLLGPLNMDALDVALRALBORHETLRTTTEAQKDIGVQVVEAGM 9717
Qy 82 ----- 81
Db 9718 KRLKVLDSLSDKNEKHEMAVLENEQMRPFTLASRPGMKHGLARLGPTEYILSLVMHMFSD 9777
Qy 82 GWS ----- 84
|||
Db 9778 GWSVDILRQELGQFYSAALRGDPLSQVKPLPIQYRDPFAAQKEAAQVAEHEROLAYWEN 9837
Qy 85 ----- 84
Db 9838 QLADSTPGELLTDPPRPOFLSGKAGVIPVTIEGPPVYKLLKFSKERQVTLFSLVLLTAPRA 9897
Qy 85 ----- 84
Db 9898 THFRLTGAEDATIGTPIANRNRPELEHIIIGFFVNTQCMRLLDGTGSTFESLVQHVRSVAT 9957
Qy 85 ----- 84
Db 9958 DAYSNQDIPFERIVSALLPGSDASRSLIQLMFALHSQDPLGNITLREGLERLPTSA 10017
Qy 85 ----- 88
OTIN
:|
Db 10018 TRFDMFHLFOEPNKLSGSILFADELFPQETINSVVTVFQILRGLDQPOVSISTMPLT 10077
Qy 89 ----- 88
Db 10078 DGLIDLEKLGLEIESSNPPRDYSVVDFVROOVAANPNAPAVVDSETSMTSLDQSEQ 10137
Qy 89 ----- 90
PE
||
Db 10138 IAAHLAQLRPESLICVMAPRSPETIVSLFGILKAGVAYLPDYNSPAARIQIPILSEVE 10197
Qy 91 ----- 90
Db 10198 GKRLVLLGSGIDMPOSMDVETARIQDILNTKVERSDPMSRPSATSLAVVIFTSGTG 10257
Qy 91 ----- 90
Db 10258 RPKGMIEHRNILLVKQSNVTSQPLQDLRMAHISNLAFAFSIWEIFTAILNGGALICID 10317
Qy 91 ----- 90
Db 10318 YFTLLDSQALRTTFEKARVNATLFPALLKECLNHAPTLFEDLVLYIGDRDLATDAK 10377
Qy 91 ----- 90
Db 10378 IQALVKGTVYNAYGPTENTVMSTIYRLTDGESYANGVPIGNAVSSGAYIMDQKORLVPP 10437
Qy 91 ----- 90
Db 10438 GVMGELVSGDGLARGYNSTLNADRFVDIVINDOKARAYRTGDRTRYPKDGSTIEFFGR 10497
Qy 91 ----- 90
Db 10498 MDQQVKIRGRHVEPAEQAMLGNKAIHDAVVVQAVDQETEMIGFVSMASDRFSEGE 10557
Qy 91 ----- 93
DDT
|||
Db 10558 EITNQVQEMEDHFEATAYAGIEAIDQATLGRDFTSWTSWYNGNLIDKAEMEELDDTQMS 10617
Qy 94 ----- 93
Db 10618 LLDKEDARPCAEIGTGTGNVLNLPKNDGLESVYVGIERSAALFVDKAAQDFPGLOKGT 10677
Qy 94 ----- 93
Db 10678 QILVGTAEIDKLKDFHPDWWVINSVAQYFPRSRYLVQIASELIHMTSVKTIFFGDMRSW 10737

Qy 94 ----- DP ----- 95
||
Db 10738 ATNRDFLVSRALYTLGDKATKDOIROEVARLENEDELLVDPAFTSLTSQWPKVKHVE 10797
Qy 96 ----- 95
Db 10798 ILPKRMTSNELSSYRYAAVLHICRDGGRNRYGRRVHSVEENAWIDFASSGMDRHALVQ 10857
Qy 96 ----- 95
Db 10858 MLDERDAKTVAIGNIPHSNTINERHTTSLDTEGEGIAQDSLGSASQATKAMAARCP 10917
Qy 96 ----- 95
Db 10918 CLSVTELVEIGAAGFRVEVSARORSQHGALDVVFHLEDDRVGRVLINFPDTERLPP 10977
Qy 96 ----- 95
Db 10978 STGLTSRPLQIRNRRFESQIREOLTLPPYVMPRSIVVLERMPLNANSKVORKELARK 11037
Qy 96 ----- 97
GH
||
Db 11038 ARTLQTIKPSATRVAPRNDIEAVLCDEFOAVLGVTVGVMDNFFELGCHSLMATKLAARLS 11097
Qy 98 ----- 101
ADLV
:|
Db 11098 RRLDTRVSVKDIENQPTILODLADVQVOTGSAPHEAIPSTPYSGPVEQSFSGRLWFLDOLN 11157
Qy 102 ----- 101
Db 11158 LNASWYHMLASRLRGLRIEALQSALATIEARHESLTTTFFESODGVVVOIVRAARNKOL 11217
Qy 102 ----- 101
Db 11218 RIIDVSGTEYLAALKQEODAAFDLTAEPGRVALLRLCPDDHVLISVMHHIISDCMSV 11277
Qy 102 ----- 103
LY
||
Db 11278 DILRQELGQLYSNASSQAPLPIQYRDFAIWQKODSIAEHOKOLNWKRLVNSKPABL 11337
Qy 104 ----- 103
Db 11338 LADTPRKALSGDADVIPIEIDDQVYONLRSFCRARHVTSFVALLAAFRAAHYRLTGAED 11397
Qy 104 ----- 103
Db 11398 ATIGSPIANRNRPELEGLICGFVNTQCLRIPVKSEDTFTLVKQARETATEAQDNQDVPP 11457
Qy 104 ----- 111
TFQLEL
:|
Db 11458 ERIVSSMVASSRDTSRNPLVQVMFAVHSQHDGLGNIRLEGVEGKPVSMASSTRFDEMHFL 11517
Qy 112 ----- 111
Db 11518 EDOGLGNGVVFVKDLFESETIRSVAVFOETLRRGLANPHANLPLTDLGSLSLSC 11577
Qy 112 ----- 113
PD
||
Db 11578 LOVNQPDYPRDASVIDVFEQVASIPKSIADASSOLTYTELDERSSOLATWLRQVTV 11637
Qy 114 ----- 118
GNRQV
:|
Db 11638 PEELGVGLAPRSCETIIAFLGIKANLAYLPDYNAPAGRIETILSSLPGNRLILGSDT 11697
Qy 119 ----- 119
R
||
Db 11698 QAVKLHANSVRFRTRISDALVESGPTEELSTRPTAOSLAYVMFTSGTGVKGMVMEHR 11757
Qy 120 GVT ----- 122
:|
Db 11758 GITRLVKNSNVAKQAAAAIAHLSNIAFDASSWEIYAPLLNGTGVVCIIDYTTIDIKAL 11817

QY 123 ----- 122
Db 11818 EAVFKOHHRGAMLPALLKQCLVSAPTWISSLEILFAAGDRLLSSQDAILARRAVSGVY 11877
QY 123 ----- 122
Db 11878 NAYGPTENTVLSITHIGENEAFNGVPIGNVSNAGFVMDQNOQLVSAGVIGELVWTG 11937
QY 123 ----- 122
Db 11938 DGLARGYTDKLRVDRFIYITLDGNVRAYRTGDRVHRPKDQIEFFORMDQOIKIRGH 11997
QY 123 ----- 122
Db 11998 RIEPAEVEQALARDPAISDAVITQLTDEEPELVAFFSLKGNANGTNGVSDQEKID 12057
QY 123 ----- 122
Db 12058 GDEQHALLMENKIRHNLQALLPTYMIPSRIIHVDQLPVNANGKIDRNELAVRAQATPTS 12117
QY 123 ----- 122
Db 12118 SVSTYVAPRNDIETIIEKEFADILSVRVGITDNFFDLGGHSLIATKLAARLSRLDTRVS 12177
QY 123 ----- 122
Db 12178 VRDVFDPVVGQJAAASIQGSTPHEAIPALSHSGPVQOQFAQGRLLWFLDRFNLAAYIM 12237
QY 123 ----- 122
Db 12238 PFCVRLRGPLRVDAQTALRALREHERHELLRTTTEEQDQGMQIVHSPMRDICVWDISGA 12297
QY 123 ----- 122
Db 12298 NEDLAKLKEEQAPFNLSDEVAMRVALFKAGENHILSIVMHHIISDGWSVDIFQOELAQ 12357
QY 123 ----- 122
Db 12358 FYSVAVRGHPDLSQVRPLPIHYRDFAVWQRQKQAVHESQLQYWIEQLADSTPAEILSD 12417
QY 123 ----- 122
Db 12418 FNRPEVLSGEAGTVPIVIEDEVYKLSLFCRNHQVTSFVLLAAFRVAHYRLTGAEDATI 12477
QY 123 ----- 122
Db 12478 GTPIANRNRPELEDLIGFFVNTQCMRIALBEHDNFLSVVRRVRSTAASAFENQDVPPERL 12537
QY 123 ----- 122
Db 12538 VSALLPGSRDASRNPVLVQMFVHVSQRNLKQLEGLEGEPTPYTATTREDFVEFHLFEQD 12597
QY 123 ----- 122
Db 12598 KGLAGNVFAADLFEAATIRSVVEVPHEILRRGLDQPDIAISTMPLVDGLAALNSRLPA 12657
QY 123 ----- 122
Db 12658 VEDIEPDFATEASVVDVFQVWNPDALAVTDTSTKLTYAELDQOSDHDVAWLSKQXLP 12717
QY 123 ----- 122
Db 12718 AESIVVVLA PRSSETIVACIGILKANLAYLPMDNSVPEARQAILSEIPGEKFVLLGAGV 12777
QY 123 ----- 122
Db 12778 PIPDNKTADVWVFI SDIVASKTDKSYSGTPRPSASSLAYVIFTSGSTGRPKGMVBHRG 12837
QY 123 ----- 122
Db 12838 VISLVKQNASRIPQSLRMAHVSNIAPDASVMEIFTTLLNGTGLFCISYFTVLDSKALSAA 12897
QY 123 ----- 122

Db 12898 FSDHRINITLPPALLKQCLADAPSVLSSLESYIGGDRLDGADATKVKDLVKGKAYNAY 12957
QY 123 ----- 122
Db 12958 GPTENSVMSTIYTIETHEFTFANGVPFGTSLGPKSKAYIMDQOQLVPAGVMGELVVGDL 13017
QY 123 ----- 122
Db 13018 ARGYTDPSLNTGRFIHITIDGKQVQAYRTGDRVYRPDYQIEFFGRLDQOIKIRGHRIE 13077
QY 123 ----- 122
Db 13078 PAEVEQALLSDSSINDAVVSAQNKEGLEVMGYITTOAAQSVDEEASNKVQEWAEHAFDS 13137
QY 123 ----- 122
Db 13138 TAYANIGGIDRDALGQDFLSWTSMYDGLIPREEMQEWLNDTMRSLLDNQPPGKVEIGT 13197
QY 123 ----- 122
Db 13198 GTGMVLFNLKVEGLQSVAGLEPSRSVTAWYNKALETFFPSLAGSARVHVGTAEIDISSIDG 13257
QY 123 ----- 122
Db 13258 LRSDLWINSVAQYPPSREYLAELTANLIRLPGVKRIFFGDMRTYATNKDFLVARAVHTL 13317
QY 123 ----- 122
Db 13318 GSNASKAMVROQVAKLEDBELLVDPAFFTSLSDQFPDEIKHVEILPKRMAANTNELSSY 13377
QY 123 ----- 122
Db 13378 RYAAVHYGGHQPNGEDEDKQWAVKIDINPKAWVDFACTMRDRQALLQLLQDRQRQDDVV 13437
QY 123 ----- 122
Db 13438 AVSNIPYKTIMERHLSQSDDEDEGTSVADGTAWISRTOSRAKECPALSADVADLIEIGKG 13497
QY 123 ----- 122
Db 13498 IGFEVEASWARQHSORGGLDAVFRHFPFPRHSGHVMFPFTEHKGRSSSLTNRPJHLLQ 13557
QY 123 ----- 122
Db 13558 SRRLEAKVRERLQSLPPYMIPSRITLLDQMLTNSGVKDRKKLARQARVIPRASAATLD 13617
QY 123 ----- 126
Db 13618 FVAPRTEIEVLCEBFTDLLGVKVGITDNFFELGCHSLATKLSARLSRRQDAGITVKOV 13677
QY 127 ----- 126
Db 13678 FDQPVADLAASILQSSSRHSRISPLPYEGPVEQFAOGRLLWFLDQFNIDALWYLI PPAL 13737
QY 127 ----- 126
Db 13738 RMRGPLOYDALAAALVALEERHESLRTTTEBRDGVGVIOVQPLRTTKDIRIIDVSGMRD 13797
QY 127 ----- 132
Db 13798 DAYLEPQKEOQTPFDLASEPGMRVALLKLGKDDHILSIVMHHIISDGWSTEVLORELGO 13857
QY 133 ----- 132
Db 13858 FYLAASKGAPLSQVAPLPIQYRDFAVWQREQVAESORQLDYWKQLADSSPAELLAD 13917
QY 133 ----- 132
Db 13918 YTRPNVLSGEAGSVFVINDSVYKSLVFCRSRQVTTFTLLAAPRAAHYRMTGSDATI 13977
QY 133 ----- 136
| :
SCL

Db	13978	GTPIANRRPELENLIGCFVNTQCMRITIGDDETFESLVQVQRSTTATAFENQDVPFERI	14037
Qy	137	-----TETGFDL-----	144
Db	14038	VSTLSAGSRDTSRNPLVOLLFAVHSQQGLGRIQLDGVVDEPVLSTVSTRFDFEHAQEA	14097
Qy	145	-----	144
Db	14098	DRLNGSYMFATDLFQPETIQGFVAVVEVLQGLEQPOSPIATWPLAEGIAQLRDAGALQ	14157
Qy	145	-----	144
Db	14158	MPKSDYPNRSASLVDFQOQAMASPSTVAVTDSTSKLTVAELDRLDQAASYLRRQQLPAE	14217
Qy	145	-----	144
Db	14218	TMVAVLAPRSCETIIAFLAILKANLAYMPLDVNTPFSARMEAIISVPGRRLLLYGSGVRH	14277
Qy	145	-----	144
Db	14278	ADINVPNAKTMILSDTVTGTDAIGTPEPLVVRPSATSLAYVITSGSGTKPKGMVEHRA	14337
Qy	145	-----	144
Db	14338	IMRLVKDSNVVTHMPATRMHAHVNIAPFDSVLFEMCATLLNGTLVCIDYLLDSTMLR	14397
Qy	145	-----	144
Db	14398	ETFEREQVRAAIFPPALLURQCLVNMPPDAIGMLEAVYVAGDRFHSRDRATOALAGPRVYN	14457
Qy	145	-----	144
Db	14458	AYGPTENAILSTIYNIDKHPYNGVPIGSAVNSGAYVMDRNQOLLPPGVMGELVWTGE	14517
Qy	145	-----	144
Db	14518	GVARGYTDASLDTDRFVTITDGORQRAYRTGDRVYRPRKGFQIEFGRLDQQAIRGHR	14577
Qy	145	-----	144
Db	14578	VELGEVEHALLSENSVTDAAVVLRTMEEEDPOLVAFTVTTDHEYRSGSSNEEDPYATQAA	14637
Qy	145	-----	144
Db	14638	GDMRKLRLSLLPYMYVPVSRTILRQMLPNANGKVDKDLARRAQMTPTASSSGPVHVAPR	14697
Qy	145	-----GVT 147.	
Db	14698	NETEAAICDEFETILGVKVGIT 14719	

RESULT 12

RESULT I
AAB41379

AAB41379
 ID AAB41379 standard; Protein; 1784 AA.

XX
XX

AC AAB41379;

XX
XX

DT	08-FEB-2001 (first entry)
XX	
DE	Human ORFX ORFL143 polypeptide sequence SEQ ID NO: 2286.
XX	
KW	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW	vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW	anticonvulsant; osteopathic; aniarthritic; immunosuppressant; cardiac;
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW	antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW	cholesterol ester storage; systemic lupus erythematosus; infection;
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn wound;

bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.

Homo sapiens.

WO200058473-A2

05-OCT-2000.

31-MAR-2000: 2000WO-US08621.

31-MAP-1999. 99118-0127607

31-MAR-1999; 99US-0127607;
02-APR-1999; 99US-0127636;

02-APR-1999; 3303-0127030;
05-APR-1999; 99US-0127728;

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shimkets RA. Leach M:

WPT: 2000-602362/57.

WFI; 2000-60236Z/
N-PSDB: AAC75588

Novel nucleic acids and peptides derived from open reading frame X useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease-

Claim 11: Page 1678-1682: 5507pp: English:

AAC74446 to AAC77606 encode the proteins given in ABA40237 to ABA43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulvovaginitis; antiparkinsonian; nootropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 1784 AA;

ery Match 30.8% Score 238; DB 21; Length 1784;

1st Local Similarity 5.6%; Pred. No. 0.038;

Matches	70;	Conservative	25;	Mismatches	51;	Indels	1067;	Gaps	13;
---------	-----	--------------	-----	------------	-----	--------	-------	------	-----

1 AAGGILHLEL-----LVAVG----- 15

57 AGGSFLSYELWPRALRKRDVSVRDAPAFYELQYRGRELRFNLTANQHLLAPGFVSETRR 116

1. **Introduction**
 2. **Background**
 3. **Methodology**
 4. **Results**
 5. **Discussion**
 6. **Conclusion**
 7. **References**
 8. **Appendix**
 9. **Figure 1**
 10. **Figure 2**
 11. **Figure 3**
 12. **Figure 4**
 13. **Figure 5**
 14. **Figure 6**
 15. **Figure 7**
 16. **Figure 8**
 17. **Figure 9**
 18. **Figure 10**
 19. **Figure 11**
 20. **Figure 12**
 21. **Figure 13**
 22. **Figure 14**
 23. **Figure 15**
 24. **Figure 16**
 25. **Figure 17**
 26. **Figure 18**
 27. **Figure 19**
 28. **Figure 20**
 29. **Figure 21**
 30. **Figure 22**
 31. **Figure 23**
 32. **Figure 24**
 33. **Figure 25**
 34. **Figure 26**
 35. **Figure 27**
 36. **Figure 28**
 37. **Figure 29**
 38. **Figure 30**
 39. **Figure 31**
 40. **Figure 32**
 41. **Figure 33**
 42. **Figure 34**
 43. **Figure 35**
 44. **Figure 36**
 45. **Figure 37**
 46. **Figure 38**
 47. **Figure 39**
 48. **Figure 40**
 49. **Figure 41**
 50. **Figure 42**
 51. **Figure 43**
 52. **Figure 44**
 53. **Figure 45**
 54. **Figure 46**
 55. **Figure 47**
 56. **Figure 48**
 57. **Figure 49**
 58. **Figure 50**
 59. **Figure 51**
 60. **Figure 52**
 61. **Figure 53**
 62. **Figure 54**
 63. **Figure 55**
 64. **Figure 56**
 65. **Figure 57**
 66. **Figure 58**
 67. **Figure 59**
 68. **Figure 60**
 69. **Figure 61**
 70. **Figure 62**
 71. **Figure 63**
 72. **Figure 64**
 73. **Figure 65**
 74. **Figure 66**
 75. **Figure 67**
 76. **Figure 68**
 77. **Figure 69**
 78. **Figure 70**
 79. **Figure 71**
 80. **Figure 72**
 81. **Figure 73**
 82. **Figure 74**
 83. **Figure 75**
 84. **Figure 76**
 85. **Figure 77**
 86. **Figure 78**
 87. **Figure 79**
 88. **Figure 80**
 89. **Figure 81**
 90. **Figure 82**
 91. **Figure 83**
 92. **Figure 84**
 93. **Figure 85**
 94. **Figure 86**
 95. **Figure 87**
 96. **Figure 88**
 97. **Figure 89**
 98. **Figure 90**
 99. **Figure 91**
 100. **Figure 92**
 101. **Figure 93**
 102. **Figure 94**
 103. **Figure 95**
 104. **Figure 96**
 105. **Figure 97**
 106. **Figure 98**
 107. **Figure 99**
 108. **Figure 100**
 109. **Figure 101**
 110. **Figure 102**
 111. **Figure 103**
 112. **Figure 104**
 113. **Figure 105**
 114. **Figure 106**
 115. **Figure 107**
 116. **Figure 108**
 117. **Figure 109**
 118. **Figure 110**
 119. **Figure 111**
 120. **Figure 112**
 121. **Figure 113**
 122. **Figure 114**
 123. **Figure 115**
 124. **Figure 116**
 125. **Figure 117**
 126. **Figure 118**
 127. **Figure 119**
 128. **Figure 120**
 129. **Figure 121**
 130. **Figure 122**
 131. **Figure 123**
 132. **Figure 124**
 133. **Figure 125**
 134. **Figure 126**
 135. **Figure 127**
 136. **Figure 128**
 137. **Figure 129**
 138. **Figure 130**
 139. **Figure 131**
 140. **Figure 132**
 141. **Figure 133**
 142. **Figure 134**
 143. **Figure 135**
 144. **Figure 136**
 145. **Figure 137**
 146. **Figure 138**
 147. **Figure 139**
 148. **Figure 140**
 149. **Figure 141**
 150. **Figure 142**
 151. **Figure 143**
 152. **Figure 144**
 153. **Figure 145**
 154. **Figure 146**
 155. **Figure 147**
 156. **Figure 148**
 157. **Figure 149**
 158. **Figure 150**
 159. **Figure 151**
 160. **Figure 152**
 161. **Figure 153**
 162. **Figure 154**
 163. **Figure 155**
 164. **Figure 156**
 165. **Figure 157**
 166. **Figure 158**
 167. **Figure 159**
 168. **Figure 160**
 169. **Figure 161**
 170. **Figure 162**
 171. **Figure 163**
 172. **Figure 164**
 173. **Figure 165**
 174. **Figure 166**
 175. **Figure 167**
 176. **Figure 168**
 177. **Figure 169**
 178. **Figure 170**
 179. **Figure 171**
 180. **Figure 172**
 181. **Figure 173**
 182. **Figure 174**
 183. **Figure 175**
 184. **Figure 176**
 185. **Figure 177**
 186. **Figure 178**
 187. **Figure 179**
 188. **Figure 180**
 189. **Figure 181**
 190. **Figure 182**
 191. **Figure 183**
 192. **Figure 184**
 193. **Figure 185**
 194. **Figure 186**
 195. **Figure 187**
 196. **Figure 188**
 197. **Figure 189**
 198. **Figure 190**
 199. **Figure 191**
 200. **Figure 192**
 201. **Figure 193**
 202. **Figure 194**
 203. **Figure 195**
 204. **Figure 196**
 205. **Figure 197**
 206. **Figure 198**
 207. **Figure 199**
 208. **Figure 200**
 209. **Figure 201**
 210. **Figure 202**
 211. **Figure 203**
 212. **Figure 204**
 213. **Figure 205**
 214. **Figure 206**
 215. **Figure 207**
 216. **Figure 208**
 217. **Figure 209**

```

16  - - - - - PD - - - - - VFQAHQED - - - - - 25
      |         |         |         |         |

```

117 PCCICBAHHTBACHII:GEVONDEI.EGGI.AA.TSACDGLKGVFOLSNEYFIEPLDSAP 176

117 RGGLGRAHIRAHITPACHLLGGEVQDPELEGGGLAAISACDGLRGVFCESNEDIFIEFDOSAF 170

26 ----- 25

QY 57 MYLITEGAPNITANLTSSLLSVCGWSOTINPEDDTDPGHADLVL----- 102
Db 297 LVLLSEDEEDLKITHAONTLASFCKWQKSNKNGDAHLPHDHTAILLTRKDLCAAMNRP 356
QY 103 ----- 102
Db 357 CETLGLSHVAGNCQPHRSCSINEDTGLPLAFTVAHELHGSFGIOHDSGNDCEPVGRKPF 416
QY 103 -----YITRF----- 107
Db 417 INSPOLLYDAAPLTWRSRCSROYITRFLDRGWGLCLDDPPAKDIIIDFPSPVPGVLYDVSHQ 476
QY 108 ----- 107
Db 477 CRLOYGAYSACFEDMDNVCHTLWCSVGTTCCHSKLDAVDGTRCGENKWCLSGECVPVGR 536
QY 108 ----- 107
Db 537 PEAVDGSWSGSAWSICSRSCGWSQSAERQCTOPTPKYKGRYCVGERKRFRLCNLAQCP 596
QY 108 ----- 107
Db 597 AGRPSPRHVQSHFDAMLYKGOLHTWVNVVNDVNPCELHCRPANEFYFAKKLRDAVVDGTP 656
QY 108 -----DLE----- 110
Db 657 CYQVRASRDLCINGICKNVGCFEIDSGAMEDRCGVCHGNGSTCHTVSTFXRRPRVXYG 716
QY 111 -----LPDGNRQVR----- 119
Db 717 VDVLIPAGAREIRIQEVAEAAFLARSEDPEKYFLNGWTIOMNGDIQVAGTFTYAR 776
QY 120 -----GYTQ--- 123
Db 777 RGNWNLTPGPTKEPWIQVPASRPGGSGRGVPRPPTLHGRSPRGVSGSVTEPGS 836
QY 124 -----LGG----- 126
Db 837 EPGPPAAASTSVPSLKNPVLVAHVHGGWQAPLGLGGWRRHLVLMGRPPLTQLLFQES 896
QY 127 ----- 126
Db 897 NPGVHYEYTIHREAGGHDEVPPVFSWYHGPMTKCTVTCRGVQRQNVYCLERQAGPVDE 956
QY 127 -----ACSP----- 130
Db 957 EHCPLGRPDDQQRKCEQPCPARWAGEMQLCSCGPGLSRRRAVLCIRSGLDEQSA 1016
QY 131 -----TW-----SCLITDGTG--- 142
Db 1017 LEPPACEHLPRPTETPCNRHVPCPATWAGNWSQCSVTCTGCTQRNVLCITNDTGVPD 1076
QY 143 ----- 142
Db 1077 EAQPAASEVTCSLPLCRWPLTGLPEGSGSGSSHELNEADFIPLHAPRPSASSPKP 1136
QY 143 ----- 142
Db 1137 GTWGNATIEEAPELDLPGVFVDDFYVDYFNFINFHEDLSYGPSEEDDLAAGTDRTPPP 1196
QY 143 ----- 142
Db 1197 HSHPAAPSTGSPVPATEPPAAKEEVLGPHSPSPWPSQAGSRPPPEQTPGNPLINFLP 1256
QY 143 -----DLGV 146
Db 1257 EEDTPICAPDLGL 1269

RESULT 13
AAB72283
ID AAB72283 standard; Protein; 997 AA.
XX

AAAB72283;
14-MAY-2001 (first entry)
Human ADAMTS-7 amino acid sequence.
ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;
tumour cachexia; inflammation; dermatosparaxis; EDS-VIIC; angiogenesis;
Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;
metastasis; embryogenesis; egg implantation; chromosome 15; ADAMTS-7.
Homo sapiens.
WO200111074-A2.
15-FEB-2001.
03-AUG-2000; 2000WO-US21223.
06-AUG-1999; 99US-0369364.
(CLEV-) CLEVELAND CLINIC FOUND.
(APTE/) APTE S S.
(HURS/) HURSKAINEN T L.
(HIRO/) HIROHATA S.
Apte SS, Hurskainen TL, Hirohata S;
WPI; 2001-159978/16.
N-PSDB; AAF63440.
Murine and human 'A Disintegrin-like And Metalloprotease domain with
Thrombospondin type I motifs' proteins and the nucleic acids encoding
them, useful for treating e.g. tumours, inflammation and arthritis -
Claim 15; Fig 4; 181pp; English.
This invention relates to murine and human ADAMTS-N (A disintegrin-like
and metalloprotease domain with thrombospondin type I motifs) proteins,
designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the
invention are cDNA sequences encoding the proteins, and antibodies
specific for the proteins. The nucleic acid sequences and proteins may be
used in the prevention, diagnosis and treatment of diseases associated
with inappropriate ADAMTS-N expression. Disorders that may be treated
using the nucleic acids, proteins and antibodies include, for example
tumour cachexia, inflammation, dermatosparaxis in cattle or Ehlers-Danlos
syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage
in arthritic (both inflammatory and non-inflammatory) disease,
angiogenesis, tumour growth and metastases, and they may also be used for
controlling embryogenesis and implantation of fertilised eggs. The
present sequence represents human ADAMTS-7. The human ADAMTS-7 gene is
located on chromosome 15.
Sequence 997 AA;
Query Match 30.7%; Score 237; DB 22; Length 997;
Best Local Similarity 18.2%; Pred. NO. 0.01;
Matches 61; Conservative 24; Mismatches 62; Indels 188; Gaps 6;
QY 1 AAGGTLHLEL-----LVAVG----- 15
Db 53 AGGSFLSYELWPRALRXRDVSVRRDAPAFYLOYRGRELRLNLTANQHLLAGFVSETRR 112
QY 16 -----PD-----VFOAQED----- 25
Db 113 RGGLGRAHRAHTPACHLLGEVDPELEGLAAISACDGLKGVFQLSNEDYFIEPLDSAP 172
QY 26 ----- 25
Db 173 APRGHAQPHVYKQAPERLAQRGSSAPSTCGVQVYYPELESRRERWEORQWRRLRR 232
QY 26 -----TERYVLTNLNIGAEELRDSGLAQPFVHLV 55

Db 233 LHORSVSEKWCETLVVADAKMVEYHGQPOVESYVLTIMMVAGLFHDPDSIGNPIHITIV 292

Qy 56 KVVLTPECAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRDLPLDCGN 115

Db 293 RLVLEDEBEDLKITHAONTLTLSFKWOKSINMKGDAPLHHDDTAILLTRKDL-CAAMN 351

Qy 116 R-QVRGVTLQGCASPTWSCLITEDTGFDLGTI 148

Db 352 RPECITGLSHVAGMCQPHRSCSINEDTGLPLAFTV 386

RESULT 14

AAU72897

ID AAU72897 standard; Protein; 1505 AA.

AC AAU72897;

DT 26-FEB-2002 (first entry)

DE Human metalloprotease partial protein sequence #9.

XX Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;

KW vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;

KW hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;

KW anorectic; antiinflammatory; aspartyl protease; cysteine protease;

KW metalloprotease; serine protease; cancer; haematopoietic; breast; colon;

KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;

KW immune-related disease; cardiovascular disease; neuronal disease;

KW migraine; sexual dysfunction; mood disorder; attention disorder;

KW cognition disorder; hypotension; hypertension; psychotic disorder;

KW dyskinesia; metabolic disorder; inflammatory disorder.

XX Homo sapiens.

OS Homo sapiens.

XX WO200181782-A2.

PN 08-NOV-2001.

XX 04-MAY-2001; 2001WO-US14431.

XX 04-MAY-2000; 2000US-201879P.

PR (SUGEN-) SUGEN INC.

XX Plowman GD, Whyte D, Sudareanam S, Manning G, Caenepeel S;

PI Payne V;

XX WPI; 2002-041502/05.

DR N-PSDB; AAS971180.

XX Novel protease polypeptide useful for screening for substances that may

PT be used to treat, e.g., cancers, immune-related diseases,

PT cardiovascular disease, migraine, pain, psychotic and inflammatory

PT disorders -

XX Claim 28; Figure 2F; 232pp; English.

PS The invention relates to an isolated, enriched, or purified protease

XX polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to

CC screen for substances (S) that may modulate its activity. Administering

CC S (which modulates protease activity in vitro) may be used to treat a

CC disease or disorder selected from cancers (e.g., of tissues, of blood or

CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,

CC brain, ovarian, bladder or kidney), immune-related diseases and

CC disorders, cardiovascular disease, brain or neuronal-associated diseases

CC (e.g., central or peripheral nervous system diseases, migraine, pain,

CC sexual dysfunction, mood disorders, attention disorders, cognition

CC disorders, hypotension, hypertension, psychotic disorders, neurological

CC disorders and dyskinesias), metabolic disorders and inflammatory

CC disorders. (I) may also be useful as a diagnostic tool for a disease or

CC disorder such as those above. AAU72876-AAU72910 represent human

CC protease amino acid sequences of the invention.

SQ Sequence 1505 AA;

Query Match 30.7%; Score 237; DB 23; Length 1505;

Best Local Similarity 5.9%; Pred. No. 0.029;

Matches 57; Conservative 30; Mismatches 51; Indels 830; Gaps 8;

Qy 3 GGIL-----

Db 149 GGLTGTGKGNGEYFLEIPMKADGNEYEDGHKPHLYIROLNNSFLOTLKYCVSVESQI 208

Qy 7 -----HLELLVAVG 15

Db 209 KETSLPFHTYSNMEDLNVMKERVLTGHTSKNVLKDERHRSKKRLISYPRVIEIMVTAD 268

Qy 16 PDVFAQHEDTERVYVLTNLNIGAEALLRDPGLGAOPRVHLVKMVLITEPEGAPNITANLTS 75

Db 269 AKVSAHGSNLQNYILTLMSIVATYIKPSIGNLHIHVVLVMTHREEGVPINFDGAT 328

Qy 76 SLLSVCGWSQTINPEDDTPGHADLVLYITRDL----- 109

Db 329 TLKNCFSWOOTONDLDDVHPSHHDTAVLITREDICSSKEKCNMLGLSVLTICDPLQSCF 388

Qy 110 ----- 109

Db 389 INEEKGLISAFTHAELGHTLGVQHDDNPRCKEMKVTVHVMAPALSPHMSPSWNSCSR 448

Qy 110 -----ELP-----DGNRQ----- 117

Db 449 KYVTBFLDTGYCECLLDKPDDEIYNLPSELPSRVDGNKQCELAFGPGSQMCPHIENICM 508

Qy 118 ----- 117

Db 509 HLWCTSTEKLHKGCTQHVPPADGTCGPMHCRHGLCVNKETETRPVNGEWGMPWEYSS 568

Qy 118 ----- 117

Db 569 CSRTCGGIESATRCNRPENRPNNGNYCVRRMKFRSCNTDSCPKGTQDFREKQCSDFNG 628

Qy 118 ----- 117

Db 629 KHLDISGIPSNVRLPRYSGICTKDRCKLYCQVAGTNYFYLLKMDVEDGTGCTETHDIC 688

Qy 118 ----- 117

Db 689 VQGQCMAGCDHVLNSSAKIDKCGVCGDNDSSCKTITGVFNSSHYGVNVVVKIPAGATNV 748

Qy 118 -----VRGVTLQGA----- 127

Db 749 DIRQVSYSCQPDSDSYALSDAEGNLFNGNFFLLSTSKKEINVQTRTVIEYSGSNNAVER 808

Qy 128 ----- 127

Db 809 INSTNRQEKELILQVLCVGNLYNPDVHYSFNIPLEERSDMFTWDPYGPWEGCTKMCQGLQ 868

Qy 128 ----- 127

Db 869 RRNITCIHKSDHSVSDKEDHLPLPSFVTQSCNTDCELRHWHVIGKSECSOCCGYRTL 928

Qy 128 -----CS----- 129

Db 929 DIHKMYSIHGOTVQVDHDCDQLKPTQBELCHGNCVFTTRWHYSEWSQCSRSCCGGER 988

Qy 130 -----PTWS-----CLIT----- 137

Db 989 SRESYCMNFGHRLADNECQELSRVTRENCPEFCPSWAASESECLVTCGKTQKORVM 1048

Qy 138 -----EDT 140

Db 1049 CQLNVDLSDGFCNSSTKPELSLSPCELHPCASQWVGMPGCTTTTCGHGYQMRDVKCVNEL 1108

Qy 138 -----EDT 140

Db 1109 ASAVLEDT 1116

Job time : 70 secs

```
RESULT 15
AAB74944
ID AAB74944 standard; Protein; 1686 AA.
XX
AC AAB74944;
XX
DT 02-JUL-2001 (first entry)
XX
DE Human ADAM type metal protease MDTs1 protein SEQ ID NO:1.
XX
KW Human; a disintegrin and metalloprotease type metal protease; MDTs1;
KW MDTs2; MDTs3; ADAM type metal protease; cytosolic; antiarthritic;
KW cancer; arthritis; arthrosis; deformans.
XX
OS Homo sapiens.
XX
PN JP2001008687-A.
XX
PD 16-JAN-2001.
XX
PF 25-JUN-1999; 99JP-0180973.
XX
PR 25-JUN-1999; 99JP-0180973.
XX
PA (YAMA) YAMANOUCHI PHARM CO LTD.
XX
DR WPI; 2001-285362/30.
DR N-PSDB; AAF82149.
XX
PT New metal protease and metal protease gene, for use as a drug for
PT cancers, arthritis and arthrosis deformans.
XX
PS Claim 1; Page 12-17; 31pp; Japanese.
XX
CC The present sequence represents a disintegrin and metalloprotease (ADAM)
CC type metal protease designated MDTs1, isolated from human. MDTs proteins
CC have cytosolic and antiarthritic activities. They can be used as a drug
CC for cancers, arthritis and arthrosis deformans.
XX
SQ Sequence 1686 AA;

Query Match 30.7%; Score 237; DB 22; Length 1686;
Best Local Similarity 18.2%; Fred. No. 0.036;
Matches 61; Conservative 24; Mismatches 62; Indels 198; Gaps 6;

QY 1 AAGGILHLEL-----LVAVG----- 15
DB 53 AGGSFSLYELWPRALRKRDVSVRDAPAFYELQYRGRELNFNLTAHQHLLAPGFVSETRR 112
QY 16 -----PD-----VFOAQHED----- 25
DB 113 RGLGRAHRAHTPACHLLGEVDPELEGLAAISACDGLKGVFQLSNEDYFIEPLDSAP 172
QY 26 ----- 25
DB 173 ARPQHAQPHVYVKQAPERLAQRGDSAPSTCGVQVYPELESRRERWEQOQWRRLRR 232
QY 26 -----TERYVLTNINIGAEILLRDPISLGAQFRVHLV 55
DB 233 LHQRSVSKKVVETLVADAKMVEYHGQVQVESYVLTIMNVAGLFHDPISGNPIHITIV 292
QY 56 KMWILTEPEGANITANITSSLLSCGWSQTNPEDDTPGHADLVLYITRFDLELPDGN 115
DB 293 RLVLLDEDEEDLKITHADNTLKSFKQKSIKMKGDAPLHLDTAILLTRKDL-CAAMN 351
QY 116 R--QVRGVTQLGACSPWCLITDGTGDLGVTI 148
DB 352 RPECETLGUSHVAGMCMOPHRSCSINEDTGLPLAFTV 386
```

Search completed: March 20, 2003, 12:37:50

GenCore version 5.1.4 ps 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:31:32 ; Search time 49.6216 Seconds
(without alignments)
564.722 Million cell updates/sec

Title: SEQID_15
Perfect score: 716
Sequence: 1 AVGPVFOAQHEDTERVLT.....SPTWSCLITGTGDLGVIT 136

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Lieting first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rhodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	716	100.0	1427	4 Q96L37	Q96L37 homo sapien
2	201	28.1	2165	5 Q19791	Q19791 caenorhabdi
3	194	27.1	1054	5 Q9W493	Q9W493 drosophila
4	173.5	24.2	950	4 Q8TE58	Q8TE58 homo sapien
5	160	22.3	1159	4 Q8TEY8	Q8TEY8 homo sapien
6	154.5	21.6	1223	4 Q8WX58	Q8WX58 homo sapien
7	154.5	21.6	1223	4 Q8TE55	Q8TE55 homo sapien
8	153.5	21.4	1095	4 Q8TE56	Q8TE56 homo sapien
9	147.5	20.6	1072	4 Q8TE57	Q8TE57 homo sapien
10	138.5	19.3	1207	4 Q8TE59	Q8TE59 homo sapien
11	138	19.3	1091	5 Q9W126	Q9W126 drosophila
12	137	19.1	790	5 Q8T458	Q8T458 drosophila
13	128.5	17.9	1081	4 Q8TE60	Q8TE60 homo sapien
14	128.5	17.9	1229	5 Q9VF61	Q9VF61 drosophila
15	128.5	17.9	1688	5 Q8SX60	Q8SX60 drosophila
16	112.5	15.7	872	5 Q22580	Q22580 caenorhabdi

17	85	11.9	1062	5 Q19204	Q19204 caenorhabdi
18	81.5	11.4	541	16 Q9F3J1	Q9F3J1 streptomyce
19	76	10.6	802	11 Q91Y03	Q91Y03 mus musculu
20	75.5	10.5	152	10 Q04205	Q04205 arabidopsia
21	75.5	10.5	509	5 Q9NDL4	Q9NDL4 caenorhabdi
22	75.5	10.5	509	5 Q20930	Q20930 caenorhabdi
23	74	10.3	755	5 Q9VF30	Q9VF30 drosophila
24	73.5	10.3	1444	5 Q17591	Q17591 caenorhabdi
25	72.5	10.1	355	4 Q9H6J3	Q9H6J3 homo sapien
26	72	10.1	531	4 Q9UH32	Q9UH32 homo sapien
27	72	10.1	1190	5 Q9V7E5	Q9V7E5 drosophila
28	71.5	10.0	518	10 Q9LJ22	Q9LJ22 oryza sativ
29	71.5	10.0	723	4 Q9UF23	Q9UF23 homo sapien
30	71.5	10.0	963	4 Q9F202	Q9F202 homo sapien
31	71	9.9	236	16 Q9KNV5	Q9KNV5 vibrio chol
32	71	9.9	254	8 Q94X42	Q94X42 bemisia tab
33	71	9.9	443	16 Q9HT65	Q9HT65 pseudomonas
34	71	9.9	726	4 Q9Y416	Q9Y416 homo sapien
35	71	9.9	906	4 Q8WVS2	Q8WVS2 homo sapien
36	71	9.9	1104	4 Q9BSJ8	Q9BSJ8 homo sapien
37	70.5	9.8	227	2 Q9L8E8	Q9L8E8 vibrio harv
38	70.5	9.8	266	17 Q57843	Q57843 pyrococcus
39	70.5	9.8	275	5 Q9VP95	Q9VP95 drosophila
40	70.5	9.8	388	16 Q9RK22	Q9RK22 streptomyce
41	70.5	9.8	539	3 P78832	P78832 echizoaecch
42	70.5	9.8	593	11 Q925M0	Q925M0 mus musculu
43	70.5	9.8	785	16 Q98KF6	Q98KF6 rhizobium l
44	70.5	9.8	793	11 Q91X21	Q91X21 mus musculu
45	70	9.8	358	2 Q9X5K8	Q9X5K8 streptomyce

ALIGNMENTS

RESULT 1

Q96L37	PRELIMINARY;	PRT; 1427 AA.
ID Q96L37		
AC Q96L37		
DT 01-DEC-2001 (TREMBLrel. 19, Created)		
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE Von Willebrand factor-cleaving protease precursor.		
GN ADAMTS13.		
OS Homo sapiens (Human).		
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX NCBI_TaxID=9606;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC TISSUE=LIVER;		
RX PubMed=11557746;		
RA Zheng X., Chung D., Takayama T.K., Majerus E.M., Sadler J.E.,		
RA Fujikawa K.;		
RT "Structure of von Willebrand Factor-cleaving Protease (ADAMTS13), a		
RT Metalloprotease Involved in Thrombotic Thrombocytopenic Purpura.";		
RL J. Biol. Chem. 276:41059-41063(2001).		
DR EMBL; AY055376; AAL17652.1; -		
DR MEROPS; M12.241; -		
DR InterPro; IPR001590; Reprolysin.		
DR InterPro; IPR000894; TSPI.		
DR InterPro; IPR000130; Zn_MTPetdase.		
DR Pfam; PF01421; Reprolysin; 1.		
DR Pfam; PF00090; tsg_1; 4.		
DR PROSITE; PS0215; ADAM_MEPRO; 1.		
DR PROSITE; PS0092; TSPI; 1.		
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.		
DR Protease; Signal.		
KW SIGNAL	1 33	POTENTIAL.
FT CHAIN	75 1427	VON WILLEBRAND FACTOR-CLEAVING PROTEASE.
SQ SEQUENCE	1427 AA; 153632 MW; EB1BC3AABCI4442 CRC64;	

Query Match 100.0%; Score 716; DB 4; Length 1427;
Best Local Similarity 100.0%; Pred. No. 2.6e-69;

```

Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AVGPVFAHQEDTTRYVLTNLNIGAEALLRDPGLSLGAQFRVHLVKVILTEPEGAPNITAN 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 87 AVGPVFAHQEDTTRYVLTNLNIGAEALLRDPGLSLGAQFRVHLVKVILTEPEGAPNITAN 146
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 61 LTSSLSVCGSQTINPEDDTPGHADLVLYITRFDELDPGNRQVRGVTOLGGACSTPW 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 147 LTSSLSVCGSQTINPEDDTPGHADLVLYITRFDELDPGNRQVRGVTOLGGACSTPW 206
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 121 SCLITEDTGFGLGVTI 136
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 207 SCLITEDTGFGLGVTI 222
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
RESULT 2
Q19791 PRELIMINARY; PRT; 2165 AA.
AC Q19791; Q27524;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F25H8.3 protein.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showlken R.,
RA Smaldon N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinscock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL; 269361; CAA93288.1; -.
DR EMBL; 269360; CAA93288.1; JOINED.
DR EMBL; 269360; CAA93287.1; -.
DR EMBL; 269361; CAA93287.1; JOINED.
DR HSSP; P15167; 1DTH.
DR MEROPS; M12.135; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_Mtpeptdse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp.1; 14.
DR SMART; SM00209; TSP1; 18.
DR PROSITE; PS00215; ADAM MEPRO; 1.
DR PROSITE; PS50092; TSP1; 6.
DR PROSITE; PS00142; ZINC PROTEASE; UNKNOWN 1.
SQ SEQUENCE 2165 AA; 244397 MW; FCC3DABAAA9C4888 CRC64;
Query Match 28.1%; Score 201; DB 5; Length 2165;
Best Local Similarity 33.6%; Pred. No. 1-2e-12;
Matches 44; Conservative 17; Mismatches 70; Indels 0; Gaps 0;
Qy 6 VFAHQEDTTRYVLTNLNIGAEALLRDPGLSLGAQFRVHLVKVILTEPEGAPNITAN 65
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 292 MYEHCGRSLEDYVLTFLSTVASIYRHSQSLRASINNVVKKLVLTENAGRPITONAAQQL 351
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Qy 66 LSVCGWSQTINPEDDTPGHADLVLYITRFDELDPGNRQVRGVTOLGGACSTPWSCLIT 125
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 352 QDFCRWQYNDPDDSSVQHHDAVAILTRDKICRSQSGKCDTGLAEELGTCMDQKSCAIL 411
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 126 EDTGFDGLGVTI 136
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 412 EDNGLSAAFTI 422
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
RESULT 3
Q9W493 PRELIMINARY; PRT; 1054 AA.
AC Q9W493;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CG4096 protein.
GN CG4096.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10711132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doul L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler K.C., Gabrielian A.E., Garg N.S., Geibart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclib J.M., Pan S.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J., Yao Q.A.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003435; AAF46065.1; -.
DR HSSP; P15167; 1ATL.
DR MEROPS; M12.231; -.
DR FlyBase; FBgn0029791; CG4096.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_Mtpeptdse.

```



```

DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR SMART; PF00090; tsp_1; 2.
DR SMART; SM00209; TSP1; 2.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSP1; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 1054 AA; 118616 MW; DC15455555CB6212 CRC64;

Query Match      27.1%; Score 194; DB 5; Length 1054;
Best Local Similarity 32.1%; Pred. No. 2.8e-12;
Matches 44; Conservative 21; Mismatches 68; Indels 4; Gaps 2;

QY 2 VGPDVFAHQEDTERYVLTNLNIGALLRDPGLGAQFRVHLVVKMVLTEPEG--APNITA 59
DB 332 VADATMSAFHRLDNGYLLTINMWSALYKDPISGNSIEIVVRIQLDEESQLQLNITQ 391
QY 60 NLTSSLLVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNRQVRGVTLGGACSP 119
DB 392 NAKQKLDRCFSQHKLNKGSEKDPHHHDVAITLIR--KNICANNMTGLANVGCMCKPK 449
QY 120 WSLCLITDGTGFDLGVTI 136
DB 450 QSCSVNEDNGIMLSHTI 466

RESULT 4
Q8TE58 PRELIMINARY; PRT; 950 AA.
AC Q8TE58;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE A disintegrin-like and metalloprotease with thrombospondin domains.
DE Metalloprotease disintegrin 15 with thrombospondin domains.
DE ADAMTS15.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteases with
RT disintegrin and thrombospondin-1 domains.;
RL Gene 283.49-62(2002).
RL EMBL; AJ315733; CAC86014.1; -.
KW Integrin; Protease.
SQ SEQUENCE 950 AA; 103286 MW; 5DFBE18285CCCC3B CRC64;

Query Match      24.2%; Score 173.5; DB 4; Length 950;
Best Local Similarity 31.0%; Pred. No. 4.3e-10;
Matches 39; Conservative 19; Mismatches 67; Indels 1; Gaps 1;

QY 10 HQEDTERYVLTNLNIGALLRDPGLGAQFRVHLVVKMVLTEPEGAPNITANLTSLLSY 69
DB 234 HGADLEHYLLTLATAARLYRHPHSILPINIVVVKLLLRDRSDGPKVTGNAALTLRNFC 293
QY 70 GWSQTINPEDDTPGHADLVLYITRFDLELPGNRQVRGVTLGGACSPWSCLITDTC 129
DB 294 AWQKLNKVSQKHPEYWDITLFTRODL-CGATTCDTLGMADVGTGCDPKRSCSVIEDDG 352
QY 130 FDLGVT 135
DB 353 LPSAFT 358

RESULT 5
Q8TEY8 PRELIMINARY; PRT; 1159 AA.
AC Q8TEY8;

```

```

DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE ADAMTS14.
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21839041; PubMed=11741898;
RA Collie A., Vandenbergh I., Thiry M., Lambert C.A., Van Beeumen J.,
RA Li S.W., Prockop D.J., Lapiere C.M., Nussgens B.V.;
RT "Cloning and Characterization of ADAMTS-14, a Novel ADAMTS Displaying
RT High Homology with ADAMTS-2 and ADAMTS-3.";
RL J. Biol. Chem. 277:5756-5766(2002).
DR EMBL; AF366351; AAL79814.1; -.
SQ SEQUENCE 1159 AA; 127336 MW; ASB130149BF7FF34 CRC64;

Query Match      22.3%; Score 160; DB 4; Length 1159;
Best Local Similarity 29.9%; Pred. No. 1.7e-08;
Matches 41; Conservative 25; Mismatches 65; Indels 6; Gaps 3;

QY 2 VGPDVFAHQ--QEDTERYVLTNLNIGALLRDPGLGAQFRVHLVVKMVLTEPEGAPNI-TA 59
DB 200 VDDSVVRFHGKSHVQNYVLTLMNIVDEIYHDESLGWHINIALVRLIMVGYRSLSLIERG 259
QY 60 NLTSSLLVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNRQVRGVTLGGACSP 119
DB 260 NPSRSLEQVCRWAHSQORQDPSHAHHHVVFLTRQDF----GPSQMGGYAPVTGWCNCHPL 315
QY 120 WSLCLITDGTGFDLGVTI 136
DB 316 RSCALNHEDGFSSAFVI 332

RESULT 6
Q8WXS8 PRELIMINARY; PRT; 1223 AA.
AC Q8WXS8;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE A disintegrin-like and metalloprotease with thrombospondin type 1
DE motif 14 precursor.
DE ADAMTS14.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638061; PubMed=1179638;
RA Boiz H., Ramirez A., von Brederlow B., Kubisch C.;
RT "Characterization of ADAMTS14, a novel member of the ADAMTS
RT metalloprotease family.";
RL Biochim. Biophys. Acta 1522:221-225(2001).
DR EMBL; AF358666; AAL40229.1; -.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF00090; TSP1; 4.
DR Pfam; PF00090; tsp_1; 4.
DR SMART; SM00209; TSP1; 4.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSP1; 1.
DR Signal; Integrin; Protease; Metalloprotease.
FT SIGNAL 1 22 POTENTIAL.
SQ SEQUENCE 1223 AA; 133871 MW; 36394AC4D92F170F CRC64;

Query Match      21.6%; Score 154.5; DB 4; Length 1223;

```



```

RESULT 10
Q8TE59 PRELIMINARY; PRT; 1207 AA.
AC Q8TE59;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE ADAMTS19.
GN ADAMTS19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ311904; CAC84565.1; --
SQ SEQUENCE 1207 AA; 134061 MW; AF36F6BF5886FDE2 CRC64;

Query Match 19.3%; Score 138.5; DB 4; Length 1207;
Best Local Similarity 30.1%; Pred. No. 4e-06;
Matches 44; Conservative 19; Mismatches 70; Indels 13; Gaps 4;

QY 4 PDVFOAHQED-TERYVLTNLNIGAEILLRDPGLGQAFRVHLVKVMVLTPEGAPNITANLT 62
DB 335 PAMVSYHGADAARRFILTILNMVFNLFQKSLGVQVNLRIKILLIHETPELYIGHGE 394

QY 63 SLLSVCGW-----SQTINPEDDTPGH-----ADLVLYITFDLLEL-PDGNRQVRGVT 110
DB 395 KMLESFCKQHEBFEGKKNDIHLESTINWGEDMTSVDAAILITRKDFCVHKDEPDCVTGIA 454

QY 111 QLGGACSPWTSCLITEDTGFDLGVTI 136
DB 455 YLSGMCSEKRCIIAEDNGLNLAFTI 480

RESULT 11
Q9W126 PRELIMINARY; PRT; 1091 AA.
AC Q9W126;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE CG3622 protein.
GN CG3622.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherter S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.H., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Shandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Butris K.C., Bueam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

```

```

Dodon K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Foaler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith T.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003458; AAF46905.1; --
DR FlyBase; FBGN0034778; CG3622.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000130; Zn_MTPetpdae.
DR Pfam; PF01421; Reprolysin; 1.
DR PROSITE; PS50215; ADAM_MEPHO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 1091 AA; 124074 MW; 5B9673B2F12565BE CRC64;

Query Match 19.3%; Score 138; DB 5; Length 1091;
Best Local Similarity 31.0%; Pred. No. 4e-06; 57; Indels 18; Gaps 8;
Matches 45; Conservative 25; Mismatches 25;

QY 2 VGPDVFOAHQED-TERYVLTNLNIGAEILLRDPGLGQAFRVHLVKVMVLT-- 49
DB 294 VPSDLY-AHWKQNFPTNTESKVSFLAMIN-GVQLLYHHPTLGRINPVLKRLKWSW 351

QY 50 EPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITFDLLELPD---GNRQV 106
DB 352 DPPGLVR-SRDVENYLSFKQKLNPFSDADPLHYDHALVLTGLDLYTDGKANSQV 410

QY 107 RGVTLGGACSPWTSCLITEDTGF 131
DB 411 VGMATVKGMCSTSIYSCTINEAKHFE 435

RESULT 12
Q8T458 PRELIMINARY; PRT; 790 AA.
AC Q8T458;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE AT15733P.
GN CG3622.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Drengel D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarino H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:31:05 ; Search time 70.7568 Seconds
(without alignments)
256.118 Million cell updates/sec

Title: SEQID_15

Perfect score: 716

Sequence: 1 AVGPDPFOAHQEDTERVLT.....SPTWSCLITDGTGDLGVTI 136

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	716	100.0	148	23	AAE24448 Human Von Willebra
2	716	100.0	242	23	AAU97641 Human aggrecanase
3	716	100.0	933	23	AAU79217 Human ADAM-TS-like
4	716	100.0	1353	23	AAE24449 Human Von Willebra
5	716	100.0	1416	23	ABB04153 Human ADAMTS-M pol
6	716	100.0	1427	23	AAE24450 Human Von Willebra
7	713	99.6	203	22	AAG63826 Amino acid sequenc
8	713	99.6	1120	22	AAG63829 Amino acid sequenc
9	529	73.9	118	21	AAE24668 Human ORFX ORF2432
10	212.5	29.7	997	22	AAB72283 Human ADAMTS-7 ami

11	212.5	29.7	1686	22	AAB74944 Human ADAM type me
12	212.5	29.7	1686	22	AAE00913 Human 27875 ADAM-T
13	212.5	29.7	1686	22	AAE00934 Human 27875 ADAM-T
14	212.5	29.7	1784	21	AAE41379 Human ORFX ORF1143
15	210	29.3	1505	23	AAU72897 Human metalloprote
16	206.5	28.8	1690	22	AAE86949 Human metalloprote
17	201	28.1	2150	21	AAE53898 Amino acid sequenc
18	201	28.1	2165	22	AAE90617 Human secreted pro
19	194	27.1	1054	22	ABE06410 Drosophila melanog
20	189.5	26.5	1073	21	AAE21264 Human metalloprote
21	187.5	26.2	947	22	AAE86950 Human metalloprote
22	187.5	26.2	1602	23	ABG30702 Human aggrecanase
23	187.5	26.2	1629	23	ABG30703 Human aggrecanase
24	187.5	26.2	1629	23	AAO14448 Human ADAMTS-Si pr
25	187.5	26.2	1916	23	AAE19173 Human protease, PR
26	187.5	26.2	1935	23	AAU72896 Human metalloprote
27	184	25.7	1882	22	AAE72286 Human ADAMTS-9 ami
28	184	25.7	1934	22	AAE72301 Human ADAMTS-9 alt
29	178.5	24.9	727	20	AAE78435 Human ADAMTS-1 pro
30	178.5	24.9	950	20	AAE49501 Human METH1 protei
31	178.5	24.9	950	22	AAE73549 Human ADAM-type me
32	178.5	24.9	950	22	AAE50002 Human METH1. Homo
33	178.5	24.9	967	19	AAE80285 Human integrin lig
34	178.5	24.9	967	20	AAE04142 Human Tango-71 pro
35	178.5	24.9	968	22	AAE50011 Protein; SEQ ID 12
36	177.5	24.8	896	21	AAE21265 Mouse metalloprote
37	176	24.6	317	22	AAE03579 Human protease-rel
38	176	24.6	356	22	AAE03581 Human protease-rel
39	176	24.6	438	22	AAE03582 Human protease-rel
40	176	24.6	468	22	AAE03574 Human protease-rel
41	176	24.6	507	22	AAE03576 Human protease-rel
42	176	24.6	589	22	AAE03577 Human protease-rel
43	176	24.6	757	22	AAE03583 Human protease-rel
44	176	24.6	859	22	AAE72282 Human ADAMTS-6 ami
45	176	24.6	908	22	AAE03572 Human protease-rel

ALIGNMENTS

RESULT 1
AAE24448
ID AAE24448 standard; Protein; 148 AA.
XX
AC AAE24448;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human Von Willebrand factor-cleaving protease fragment #1.
XX
DE Human, Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;
KW transgenic animal; immunisation; thromboembolic disease; preclampsia;
KW thrombotic thrombocytopenic purpura; TTP; Henoch-Schonlein purpura;
KW thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
KW transgenic; anticoagulant.
OS Homo sapiens.
XX
XX WO200242441-A2.
XX
PD 30-MAY-2002.
XX
XX 20-NOV-2001; 2001WO-EP13391.
XX
XX 22-NOV-2000; 2000US-0721254;
PR 12-APR-2001; 2001US-083328;
XX
XX (BAXT) BAXTER AG.
XX
XX Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
PI Scheifflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
PI Zimmermann K, Voelkel D;
XX

DR WPI; 2002-479950/51.
 DR N-PSDB; AAD39331.
 XX Novel isolated or substantially purified Von Willebrand factor-cleaving
 PT protease, useful for producing preparation for therapy of thrombosis
 PT and thromboembolic disease such as thrombotic thrombocytic purpura -
 XX
 PS Claim 1; Fig 3; 93pp; English.
 XX
 CC The invention relates to an isolated or substantially pure Von Willebrand
 CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
 CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
 CC solution comprising vWF with the polypeptide ligand under conditions
 CC where vWF is bound to the ligand and recovering from the ligand purified
 CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
 CC which involves immunising an animal with vWF-cp and isolating the anti-
 CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
 CC producing a preparation of prophylaxis and therapy of thrombosis and
 CC thromboembolic disease such as thrombotic thrombocytic purpura (TTP),
 CC Henoch-Schönlein purpura, preclampsia, neonatal thrombocytopenia or
 CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
 CC plasmatic or recombinantly produced vWF. The invention is useful for
 CC construction expression systems and generating transgenic animals which
 CC express the polypeptide in vivo. The present sequence is human vWF-cp
 CC fragment.
 XX
 SQ Sequence 148 AA;
 Query Match 100.0%; Score 716; DB 23; Length 148;
 Best Local Similarity 100.0%; Pred. No. 3e-78;
 Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AVGPVFOAHOEDTERVYLTNLNIGAEELRDPGLGQFRVHLVKMVLTEPEGAPNITAN 60
 DB 13 AVGPVFOAHOEDTERVYLTNLNIGAEELRDPGLGQFRVHLVKMVLTEPEGAPNITAN 72
 QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRGVTQLGGACSPW 120
 DB 73 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRGVTQLGGACSPW 132
 QY 121 SCLITEDTGFDLGVTI 136
 DB 133 SCLITEDTGFDLGVTI 148
 RESULT 2
 AAU97641
 ID AAU97641 standard; Protein; 242 AA.
 AC AAU97641;
 XX
 XX 27-AUG-2002 (first entry)
 DT Human aggrecanase protein.
 DE
 XX Aggrecanase; osteopathic; antiarthritic; antiinflammatory; human;
 KW cartilage; osteoarthritis; inflammatory disease; enzyme.
 KW
 XX Homo sapiens.
 OS
 XX WO200233093-A2.
 PN
 XX 25-APR-2002.
 PD
 XX 17-OCT-2001; 2001WO-032458.
 XX
 PF 18-OCT-2000; 2000US-241469P.
 XX
 PR (GEMY) GENETICS INST INC.
 PA
 XX Racie LA, Twine NC, Agostino WJ, Wolfman NM, Morris EA;
 PI
 XX WPI; 2002-452389/48.
 XX
 DR

DR N-PSDB; ABK52579.

XX Novel purified aggrecanase polypeptide useful for developing inhibitors
 PT and antibodies to the aggrecanase polypeptide, which are useful for
 PT treating aggrecanase-associated condition such as osteoarthritis -
 XX
 PS Claim 11; Page 33-34; 41pp; English.

XX This invention relates to the cDNA and protein sequences of a novel
 CC human aggrecanase polypeptide. The protein of the invention may be
 CC used to inhibit the proteolytic activity of aggrecanase, or to inhibit
 CC the aggrecanase-mediated cleavage of aggrecan in cartilage. The
 CC protein of the invention is useful for developing inhibitors of
 CC aggrecanase protein. The cDNA sequence encoding the aggrecanase protein
 CC of the invention is useful for designing probes for obtaining DNA
 CC sequences encoding other aggrecanase molecules. The cDNA sequence is
 CC also useful for detecting mRNA encoding aggrecanase in a given cell
 CC population, and thus for detecting or diagnosing genetic disorders
 CC involving the aggrecanase, or disorders involving cellular, organ or
 CC tissue disorders in which aggrecanase is irregularly transcribed or
 CC expressed. The DNA sequences may also be useful for preparing vectors
 CC for gene therapy applications. An inhibitor of the protein is useful
 CC in treating conditions characterised by degradation of articular
 CC cartilage, by blocking the enzyme's proteolytic activity. An
 CC aggrecanase protein inhibitor and a method for inhibition of its activity
 CC are useful for treating various aggrecanase-associated conditions
 CC including osteoarthritis and other inflammatory diseases. The
 CC present sequence represents the human aggrecanase protein of the
 CC invention.

XX Sequence 242 AA;

Query Match 100.0%; Score 716; DB 23; Length 242;

Best Local Similarity 100.0%; Pred. No. 6.2e-78;

Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVGPVFOAHOEDTERVYLTNLNIGAEELRDPGLGQFRVHLVKMVLTEPEGAPNITAN 60

DB 54 AVGPVFOAHOEDTERVYLTNLNIGAEELRDPGLGQFRVHLVKMVLTEPEGAPNITAN 113

QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRGVTQLGGACSPW 120

DB 114 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRGVTQLGGACSPW 173

QY 121 SCLITEDTGFDLGVTI 136

DB 174 SCLITEDTGFDLGVTI 189

RESULT 3

AAU79217

ID AAU79217 standard; Protein; 933 AA.

AC AAU79217;

XX 15-JUL-2002 (first entry)

DT Human ADAM-TS-like protein.

DE

XX Human; ADAM-TS-like protein; cardiovascular disorder; angina;

KW vascular system; congestive heart failure; myocardial infarction;

KW ischemic heart disease; arrhythmia; hypertensive vascular disease;

KW secondary peripheral arterial hypertension; peripheral vascular disease; embolism;

KW chronic peripheral arterial occlusive disease; acute arterial thrombosis;

KW inflammatory vascular disorder; chronic obstructive pulmonary disease;

KW liver disorder.

XX Homo sapiens.

OS

XX WO200226999-A2.

PN

XX 04-APR-2002.

PD

XX

PF 26-SEP-2001; 2001WO-EP11124.
 XX
 XX 28-SEP-2000; 2000US-235881P.
 PR 25-JUL-2001; 2001US-307393P.
 XX
 XX (FARB) BAYER AG.
 XX
 XX Xiao Y;
 XX
 DR WPI: 2002-383274/41.
 DR N-PSDB; ABK49356.
 XX
 XX New purified human ADAM-TS-like protein, useful for identifying
 PT modulators of protein activity for treating cardiovascular or liver
 PT disorder or chronic obstructive pulmonary disease -
 XX
 XX Claim 25; Fig 2; 106pp; English.
 XX
 CC The invention relates to a human ADAM-TS-like protein and the
 CC polynucleotide encoding it. The protein of the invention is useful for
 CC treating cardiovascular disorders including diseases of the heart and
 CC vascular system, such as congestive heart failure, myocardial infarction,
 CC ischaemic heart diseases (e.g., stable angina, unstable angina), atrial
 CC and ventricular arrhythmia, hypertensive vascular diseases (e.g., all
 CC kinds of secondary arterial hypertension), and peripheral vascular
 CC diseases (e.g., chronic peripheral arterial occlusive disease, acute
 CC arterial thrombosis and embolism, inflammatory vascular disorders, etc),
 CC liver disorders and chronic obstructive pulmonary disease. The sequences
 CC are useful in diagnostic assays for detecting diseases and abnormalities
 CC or susceptibility to diseases and abnormalities related to the presence
 CC of mutations in the nucleic acid sequences which encode the protein. The
 CC sequences are also useful for modulating ADAM-TS-like protein activity in
 CC a disease condition. This sequence represents the human ADAM-TS-like
 CC protein.
 XX
 SQ Sequence 933 AA;
 XX
 Query Match 100.0%; Score 716; DB 23; Length 933;
 Best Local Similarity 100.0%; Pred. No. 4.4e-77;
 Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AVGPDPVFAHQEDTERVLTNINIGAEELRDPISLGAQFRVHLVKVMVILTEPEGAPNITAN 60
 DB 90 AVGPDPVFAHQEDTERVLTNINIGAEELRDPISLGAQFRVHLVKVMVILTEPEGAPNITAN 149
 QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRGVTVQLGGACSPW 120
 DB 150 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRGVTVQLGGACSPW 209
 QY 121 SCLITEDTGFDLGVTI 136
 DB 210 SCLITEDTGFDLGVTI 225
 RESULT 4
 AAE24449
 ID AAE24449 standard; Protein; 1353 AA.
 XX
 AC AAE24449;
 XX
 XX 04-OCT-2002 (first entry)
 XX
 DE Human Von Willebrand factor-cleaving protease fragment #2.
 XX
 KW Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;
 KW transgenic animal; immunisation; thromboembolic disease; preclampsia;
 KW thrombotic thrombocytopenic purpura; TTP; Henoch-Schonlein purpura;
 KW thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
 KW transgenic; anticoagulant.
 XX
 OS Homo sapiens.
 XX
 XX W0200242441-A2.
 PN

XX 30-MAY-2002.
 PD
 XX
 PF 20-NOV-2001; 2001WO-EP13391.
 XX
 XX 22-NOV-2000; 2000US-0721254.
 PR 12-APR-2001; 2001US-0833328.
 XX
 XX (BAXT) BAXTER AG.
 XX
 XX Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
 PI Scheifflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
 PI Zimmermann K, Voelkel D;
 XX
 XX WPI: 2002-479950/51.
 XX
 PT Novel isolated or substantially purified Von Willebrand factor-cleaving
 PT protease, useful for producing preparation for therapy of thrombosis
 PT and thromboembolic disease such as thrombotic thrombocytopenic purpura -
 XX
 XX Claim 1; Page 64-68; 93pp; English.
 XX
 CC The invention relates to an isolated or substantially pure Von Willebrand
 CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
 CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
 CC solution comprising vWF with the polypeptide ligand under conditions
 CC where vWF is bound to the ligand and recovering from the ligand purified
 CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
 CC which involves immunising an animal with vWF-cp and isolating the anti-
 CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
 CC producing a preparation of prophylaxis and therapy of thrombosis and
 CC thromboembolic disease such as thrombotic thrombocytopenic purpura (TTP),
 CC Henoch-Schonlein purpura, preclampsia, neonatal thrombocytopenia or
 CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
 CC plasmatic or recombinantly produced vWF. The invention is useful for
 CC construction expression systems and generating transgenic animals which
 CC express the polypeptide in vivo. The present sequence is human vWF-cp
 CC fragment.
 XX
 SQ Sequence 1353 AA;
 XX
 Query Match 100.0%; Score 716; DB 23; Length 1353;
 Best Local Similarity 100.0%; Pred. No. 7.6e-77;
 Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AVGPDPVFAHQEDTERVLTNINIGAEELRDPISLGAQFRVHLVKVMVILTEPEGAPNITAN 60
 DB 13 AVGPDPVFAHQEDTERVLTNINIGAEELRDPISLGAQFRVHLVKVMVILTEPEGAPNITAN 72
 QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRGVTVQLGGACSPW 120
 DB 73 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPDGNRQVRGVTVQLGGACSPW 132
 QY 121 SCLITEDTGFDLGVTI 136
 DB 133 SCLITEDTGFDLGVTI 148
 RESULT 5
 ABB04153
 ID ABB04153 standard; protein; 1416 AA.
 XX
 AC ABB04153;
 XX
 XX 26-MAR-2002 (first entry)
 DT Human ADAMTS-M polypeptide.
 DE
 XX Osteoarthritis; rheumatoid arthritis; inflammatory bowel disease;
 KW Crohn's disease; asthma; Alzheimer's disease; organ transplant rejection;
 KW cachexia; allergy; cancer; leukaemia; lymphoma; osteoporosis;
 KW atherosclerosis; congestive heart failure; myocardial infarction; stroke;
 KW neurodegenerative disease; autoimmune disorder; Huntington's;
 KW

KW Parkinson's; migraine; pain; depression; multiple sclerosis; burn;
KW infertility; diabetic shock; gene therapy; ADAMTS-M;
KW A Disintegrin And Metalloprotease; thrombospondin domain.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..97
FT /label= Prodomain
FT /note= "The mature form of the ADAMTS-M protein is
FT processed by furin cleavage of the prodomain"
FT 94..97
FT /label= Furin_cleavage_site
FT 98..1416
FT /label= Mature ADAMTS-M protein
FT /note= "The mature form of the ADAMTS-M protein is
FT processed by furin cleavage of the prodomain"
FT 98..311
FT /label= Metalloprotease_domain
FT 247..272
FT /label= Zinc-binding_motif
FT 324..394
FT /label= Disintegrin_domain
FT 410..473
FT /label= Thrombospondin_submotif
FT 419..424
FT /label= Heparin-binding_domain
FT 1099..1156
FT /label= Thrombospondin_submotif
XX
PN EP1152055-A1.
XX
XX 07-NOV-2001.
XX 24-APR-2001; 2001EP-0303706.
XX 27-APR-2000; 2000US-200040P.
XX (PF12) PFIZER PROD INC.
XX
XX Buckbinder L, Mitchell PG, Wachtmann TS, Walsh RT;
XX WPI: 2002-084275/12.
DR N-PSDB; ABA02549.
XX
XX New polynucleotide, useful in gene therapy, particularly for treating
PT or preventing e.g. arthritis, Crohn's disease, Alzheimer's disease and
PT organ transplant toxicity and rejection, comprises ADAMTS
PT polynucleotide and encoded polypeptide -
XX
XX Claim 4; Fig 2; 3lpp; English.
XX
XX The present sequence represents a ADAMTS protein, designated ADAMTS-M,
CC that exhibits the characteristics of the ADAM (A Disintegrin And
CC Metalloprotease) family of metalloproteases, and contains a
CC thrombospondin domain (TS). The protein is encoded by the cDNA given in
CC ABA02549. The specification describes a newly isolated polynucleotide,
CC comprising a nucleotide sequence encoding an ADAMTS-M polypeptide as
CC given in the specification, or a metalloproteinase, disintegrin domain,
CC prodomain or its thrombospondin submotif. The polynucleotide, polypeptide
CC and agent are useful for manufacturing a medicament for treating a
CC subject in need of altering activity or expression of ADAMTS-M. The
CC polynucleotide, ADAMTS-M polypeptide and agent are useful for
CC manufacturing a medicament for treating arthritis (osteoarthritis and
CC rheumatoid arthritis), inflammatory bowel disease, Crohn's disease,
CC aethma, Alzheimer's disease, organ transplant toxicity and rejection,
CC cachexia, allergy, cancer (e.g. solid tumour cancer including colon,
CC breast, lung, prostate, brain or haematopoietic malignancies including
CC leukaemia and lymphoma), osteoporosis, arteriosclerosis, aortic aneurysm,
CC congestive heart failure, myocardial infarction, stroke, head trauma,
CC spinal cord injury, neurodegenerative disease, autoimmune disorders,
CC Huntington's disease, Parkinson's disease, migraine, pain, depression,
CC multiple sclerosis, abnormal wound healing, burns, infertility or

CC diabetic shock. The polynucleotide and polypeptide are also useful for
CC diagnosing the diseases above. The polynucleotide is particularly useful
CC in gene therapy for treating the diseases cited above.
XX
SQ Sequence 1416 AA;
Query Match 100.0%; Score 716; DB 23; Length 1416;
Best Local Similarity 100.0%; Pred. No. 8.1e-77;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AVGPVFOAHOEDTERVLTNLNIGALLRDPSPSLGAQFRVHLVKMVLITPEGAPNITAN 60
DB 110 AVGPVFOAHOEDTERVLTNLNIGALLRDPSPSLGAQFRVHLVKMVLITPEGAPNITAN 169
OY 61 LTSSLLSVCGWSQINPEDDTPGHADLVLYITRFDELDPGNRQVRGVLTOLGACSPTW 120
DB 170 LTSSLLSVCGWSQINPEDDTPGHADLVLYITRFDELDPGNRQVRGVLTOLGACSPTW 229
OY 121 SCLITEDTGFDLGVTI 136
DB 230 SCLITEDTGFDLGVTI 245
RESULT 6
AAE24450
ID AAE24450 standard; Protein; 1427 AA.
AC AAE24450;
XX
XX 04-OCT-2002 (first entry)
XX Human Von Willebrand factor-cleaving protease (VWF-cp).
XX Human; Von Willebrand factor-cleaving protease; VWF-cp; therapy; enzyme;
KW transgenic animal; immunisation; thromboembolic disease; pre-eclampsia;
KW thrombotic thrombocytopenic purpura; TTP; Henoch-Schonlein purpura;
KW thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
KW transgenic; anticoagulant; chromosome 9.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Peptide 1..51
FT /label= Signal_peptide
FT 52..1427
FT /note= "Mature human VWF-cp protein"
FT Cleavage-site 69..75
FT /note= "Furin cleavage site"
FT 224..228
FT /note= "Catalytical side"
FT 249
FT /note= "Met turn"
FT 301..377
FT /note= "Disintegrin like motif"
FT 387..439
FT /note= "Thrombospondin type I motif"
FT 441..553
FT /note= "Cysteine rich region"
FT 554..687
FT /note= "Spacer"
FT 688..743
FT /note= "Thrombospondin type I motif"
FT 744..805
FT /note= "Thrombospondin type I motif"
FT 897..952
FT /note= "Thrombospondin type I motif"
FT 953..1013
FT /note= "Thrombospondin type I motif"
FT 1016..1073
FT /note= "Thrombospondin type I motif"
FT 1075..1131
FT /note= "Thrombospondin type I motif"
XX

PN WO200242441-A2.
 XX 30-MAY-2002.
 PD
 XX 20-NOV-2001; 2001WO-EP13391.
 PF
 XX 22-NOV-2000; 2000US-0721254.
 PR 12-APR-2001; 2001US-0833328.
 XX
 XX (BAXT) BAXTER AG.
 XX
 PI Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
 PI Scheiflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
 PI Zimmermann K, Voelkel D;
 XX
 DR WPI: 2002-479950/51.
 DR N-PSDB; RAD39332.
 XX
 XX Novel isolated or substantially purified Von Willebrand factor-cleaving
 PT protease, useful for producing preparation for therapy of thrombosis
 PT and thromboembolic disease such as thrombotic thrombocytic purpura
 PS
 XX Claim 1; Fig 5; 93pp; English.
 XX
 CC The invention relates to an isolated or substantially pure Von Willebrand
 CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
 CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
 CC solution comprising vWF with the polypeptide ligand under conditions
 CC where vWF is bound to the ligand and recovering from the ligand purified
 CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
 CC which involves immunising an animal with vWF-cp and isolating the anti-
 CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
 CC producing a preparation of prophylaxis and therapy of thrombosis and
 CC thromboembolic disease such as thrombotic thrombocytic purpura (TTP).
 CC Hemocho-Schonlein purpura, preeclampsia, neonatal thrombocytopenia or
 CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
 CC plasmatic or recombinantly produced vWF. The invention is useful for
 CC construction expression systems and generating transgenic animals which
 CC express the polypeptide in vivo. The present sequence is human vWF-cp
 CC protein. vWF-cp gene is located on chromosome 9.
 XX
 SQ Sequence 1427 AA;
 Query Match 100.0%; Score 716; DB 23; Length 1427;
 Best Local Similarity 100.0%; Pred. No. 8.2e-77;
 Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 AVGPDVFOAHQEDTERYVLTNLNIGALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 60
 Db 87 AVGPDVFOAHQEDTERYVLTNLNIGALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 146
 Qy 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQRQVRGVTLGGACSPW 120
 Db 147 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQRQVRGVTLGGACSPW 206
 Qy 121 SCLITDGTGFDLGVTI 136
 Db 207 SCLITDGTGFDLGVTI 222
 RESULT 7
 AAG63826
 ID AAG63826 standard; Protein; 203 AA.
 XX
 AC AAG63826;
 XX
 DT 29-OCT-2001 (first entry)
 XX
 DE Amino acid sequence of a human zdtnt5 polypeptide.
 XX Human; zdtnt5; anti-angiogenic; intestinal polypeptide; wound healing;
 KW extracellular matrix interaction; tumour suppression; gamete maturation;
 KW immunologic recognition; gastrointestinal irradiation; chemotherapy;

proteolysis; apoptosis; angiogenesis; infection; cell adhesion;
 cell fusion; cell signalling; tumour; Crohn's disease; melanoma;
 inflammatory bowel disease; food poisoning; degenerative disease;
 inflammation; fertility; gamete maturation; epithelial disorder.
 Homo sapiens.
 Key Location/Qualifiers
 Modified-site 69
 FT /note= "potential N-linked glycosylation site"
 Modified-site 73
 FT /note= "potential N-linked glycosylation site"
 Region 151..161
 FT /note= "zinc-binding motif"
 XX
 PN WO200159112-A1.
 XX
 PD 16-AUG-2001.
 XX
 XX 09-FEB-2001; 2001WO-US04198.
 PF
 XX 10-FEB-2000; 2000US-0501806.
 PR
 XX (ZYMO) ZYMOGENETICS INC.
 PA
 XX Holloway JL, Sheppard PO;
 PI WPI: 2001-522477/57.
 DR N-PSDB; AAH74759.
 XX
 XX New anti-angiogenic intestinal polypeptides, zdtnt5 polypeptides, which
 PT are members of disintegrin proteases, for modulating extracellular
 PT matrix interaction, tumour suppression and wound healing
 PS
 XX Claim 1; Page 73-74; 92pp; English.
 XX
 CC The present sequence represents a human zdtnt5 polypeptide. The zdtnt5
 CC polypeptide is an anti-angiogenic intestinal polypeptide. zdtnt5 is
 CC used for modulating extracellular matrix interactions. zdtnt5
 CC polypeptide is useful as a tool for identifying new family members
 CC of polypeptides. zdtnt5 polynucleotides are useful as probes or primers
 CC to clone 5' non-coding regions of zdtnt5 gene. zdtnt5 polypeptides are
 CC used for tumour suppression, gamete maturation, immunologic recognition,
 CC and growth and differentiation either working in isolation or in
 CC conjunction with other molecules in colon, small intestine, fetal lung,
 CC testis and B-cells. zdtnt5 polypeptides are also useful for promoting
 CC wound healing, in the treatment of disorders associated with recovery
 CC after gastrointestinal irradiation, chemotherapy or antibody use
 CC as anti-infectives, and extracellular matrix repair and remodeling. The
 CC polypeptides are also useful for modulating proteolysis, apoptosis,
 CC angiogenesis, infection, cell adhesion, cell fusion and signalling.
 CC The polypeptides are also useful for treating tumour formation, Crohn's
 CC disease, inflammatory bowel disease, food poisoning, melanoma,
 CC degenerative diseases, disorders related to immunity, inflammation,
 CC fertility, gamete maturation, immunology, trauma and epithelial
 CC disorders.
 XX
 SQ Sequence 203 AA;
 Query Match 99.6%; Score 713; DB 22; Length 203;
 Best Local Similarity 99.3%; Pred. No. 1.1e-77;
 Matches 135; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 AVGPDVFOAHQEDTERYVLTNLNIGALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 60
 Db 14 AVGPDVFOAHQEDTERYVLTNLNIGALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 73
 Qy 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQRQVRGVTLGGACSPW 120
 Db 74 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNQRQVRGVTLGGACSPW 133
 Qy 121 SCLITDGTGFDLGVTI 136
 |||||||

Dd	134	SCLITDGTGFDLGVTI 149	
AAAG63829			
ID	AAAG63829	standard; Protein; 1120 AA.	
XX	AC		
XX	AAAG63829;		
DT	29-OCT-2001	(first entry)	
XX			
DE	Amino acid sequence of a human zdint5 polypeptide.		
XX	Human; zdint5; anti-angiogenic; intestinal polypeptide; wound healing;		
KW	extracellular matrix interaction; tumour suppression; gamete maturation;		
KW	immunologic recognition; gastrointestinal irradiation; chemotherapy;		
KW	proteolysis; apoptosis; angiogenesis; infection; cell adhesion;		
KW	cell fusion; cell signalling; tumour; Crohn's disease; melanoma; trauma;		
KW	inflammatory bowel disease; food poisoning; degenerative disease;		
KW	inflammation; fertility; gamete maturation; epithelial disorder.		
XX			
OS	Homo sapiens.		
XX			
PH	Key	Location/Qualifiers	
FT	Modified-site	172 /note= "potential N-linked glycosylation site"	
FT	Modified-site	176 /note= "potential N-linked glycosylation site"	
FT	Modified-site	485 /note= "potential N-linked glycosylation site"	
FT	Modified-site	533 /note= "potential N-linked glycosylation site"	
FT	Modified-site	560 /note= "potential N-linked glycosylation site"	
FT	Modified-site	595 /note= "potential N-linked glycosylation site"	
FT	Modified-site	635 /note= "potential N-linked glycosylation site"	
FT	Misc-difference	474 /note= "unspecified residue encoded by TTN"	
XX			
PN	WO200159112-A1.		
XX			
PD	16-AUG-2001.		
XX			
PF	09-FEB-2001; 2001WO-US04198.		
XX			
PR	10-FEB-2000; 2000US-0501806.		
XX			
PA	(ZYMO) ZYMOGENETICS INC.		
XX			
PI	Holloway JL, Sheppard PO;		
XX			
DR	WPI; 2001-522477/57.		
XX	N-PSDB; AAH74765.		
XX			
PT	New anti-angiogenic intestinal polypeptides, zdint5 polypeptides, which		
PT	are members of disintegrin proteases, for modulating extracellular		
PT	matrix interaction, tumour suppression and wound healing		
XX			
PS	Claim 2; Page 84-88; 92pp; English.		
XX			
CC	The present sequence represents a human zdint5 polypeptide. The zdint5		
CC	polypeptide is an anti-angiogenic intestinal polypeptide. Zdint5 is		
CC	used for modulating extracellular matrix interactions. Zdint5		
CC	polypeptide is useful as a tool for identifying new family members		
CC	of polypeptides. Zdint5 polynucleotides are useful as probes or primers		
CC	to clone 5' non-coding regions of zdint5 gene. Zdint5 polypeptides are		
CC	used for tumour suppression, gamete maturation, immunologic recognition,		
CC	and growth and differentiation either working in isolation or in		
CC	conjunction with other molecules in colon, small intestine, fetal lung,		
CC	testis and B-cells. zdint5 polypeptides are also useful for promoting		
CC	wound healing, in the treatment of disorders associated with recovery		

CC	after gastrointestinal irradiation, chemotherapy or antibody use,		
CC	as anti-infectives, and extracellular matrix repair and remodeling. The		
CC	polypeptides are also useful for modulating proteolysis, apoptosis,		
CC	angiogenesis, infection, cell adhesion, cell fusion and signalling.		
CC	The polypeptides are also useful for treating tumour formation, Crohn's		
CC	disease, inflammatory bowel disease, food poisoning, melanoma,		
CC	degenerative diseases, disorders related to immunity, inflammation,		
CC	fertility, gamete maturation, immunology, trauma and epithelial		
CC	disorders.		
XX			
SQ	Sequence 1120 AA;		
	Query Match 99.6%; Score 713; DB 22; Length 1120;		
	Best Local Similarity 99.3%; Pred. No. 1.3e-76;		
	Matches 135; Conservative 1; Mismatches 0; Indels 0; Gaps 0;		
Oy	1	AVGPDVFOAHQEDTERVYLTNLNIGALLRDPISGAOFRVHLVKVILTEPGAPNITAN 60	
Dd	117	AVGPDVFOAHQEDTERVYLTNLNIGALLRDPISGAOFRVHLVKVILTEPGAPNITAN 176	
Oy	61	LTSSLLSVCGWSQINPEDDTPGHADLVLYITRFDLELPGNRQVRGVTOLGGACSPTW 120	
Dd	177	LTSSLLSVCGWSQINPEDDTPGHADLVLYITRFDLELPGNRQVRGVTOLGGACSPTW 236	
Oy	121	SCLITDGTGFDLGVTI 136	
Dd	237	SCLITDGTGFDLGVTI 252	
RESULT 9			
AAAB42668			
ID	AAAB42668	standard; Protein; 118 AA.	
XX			
AC	AAAB42668;		
XX			
DT	08-FEB-2001	(first entry)	
XX			
DE	Human ORFX ORF2432	polypeptide sequence SEQ ID NO:4864.	
XX			
KW	Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;		
KW	vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;		
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;		
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;		
KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;		
KW	antiviral; antibacterial; antifungal; antineumatic; antithyroid;		
KW	antianaemic; gene therapy; cancer; proliferative disorder; hypertension;		
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;		
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;		
KW	cholesterol ester storage; systemic lupus erythematosus; infection;		
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;		
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;		
KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;		
KW	thrombosis; contraceptive.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200058473-A2.		
XX			
PD	05-OCT-2000.		
XX			
PF	31-MAR-2000; 2000WO-US08621.		
XX			
PR	31-MAR-1999; 99US-0127607.		
PR	02-APR-1999; 99US-0127636.		
PR	05-APR-1999; 99US-0127728.		
PR	30-MAR-2000; 2000US-0540763.		
XX			
PA	(CURA-) CURAGEN CORP.		
XX			
PI	Shinkets RA, Leach M;		
XX			
XX	WPI; 2000-602362/57.		
DR	N-PSDB; AAC76877.		

```
XX Novel nucleic acids and peptides derived from open reading frame X.
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 4050; 5507pp; English.
XX AAC74446 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 1161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antiproliferative; antiparkinsonian; neurotropic; immunosuppressive;
CC osteoprotective; anticonvulsant; antiarthritic; vasodilator;
CC immunostimulant; cardiatic; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease, to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX SQ Sequence 118 AA;
Query Match 73.9%; Score 529; DB 21; Length 118;
Best Local Similarity 97.1%; Pred. No. 8.8e-56;
Matches 100; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 34 LGAQRVHLVKMVLTEPEGAPNITANLTSSLLSVCGWSQTINPEDTDPGHADLVLYIT 93
DB 1 LGAHRVHLVKMVLTEPEGAPNITANLTSSLLSVCGWSQTINPEDTDPGHADLVLYIT 60
OY 94 RFDLEPDGNGRQVGTOLGACSPWTSCLITDTCFDLGVTI 136
DB 61 RFDLEPDGNGXAVRGVGTOLGACSPWTSCLITDTCFDLGVTI 103
RESULT 10
AAB72283
ID AAB72283 standard; Protein; 997 AA.
XX
XX AAB72283;
XX
XX 14-MAY-2001 (first entry)
XX Human ADAMTS-7 amino acid sequence.
DE
XX ADAMTS-N; disintegrin; metalloprotease; thrombospondin type I motif;
XX tumour cachexia; inflammation; dermatosparaxis; EDS-VIIC; angiogenesis;
XX Ehlers-Danlos syndrome type VIIC; articular cartilage erosion; human;
XX metastasis; embryogenesis; egg implantation; chromosome 15; ADAMTS-7.
XX
XX Homo sapiens.
XX
XX WO200111074-A2.
XX
XX 15-FEB-2001.
XX
XX 03-AUG-2000; 2000WO-US21223.
XX
XX 06-AUG-1999; 99US-0369364.
XX
XX (CLEV-) CLEVELAND CLINIC FOUND.
PA (APTE/) APTE S. S.
PA (HURS/) HURSKAINEN T. L.
PA (HIRO/) HIROHATA S.
```

```
XX Apte SS, Hurskainen TL, Hirohata S;
XX WPI; 2001-159978/16.
XX N-PSDB; AAF63440.
XX Murine and human 'A Disintegrin-like And Metalloprotease domain with
XX Thrombospondin type I motifs' proteins and the nucleic acids encoding
XX them, useful for treating e.g. tumours, inflammation and arthritis -
XX Claim 15; Fig 4; 181pp; English.
XX This invention relates to murine and human ADAMTS-N (A disintegrin-like
XX and metalloprotease domain with thrombospondin type I motifs) proteins,
XX designated ADAMTS-5, 6, 7, 8, 9, 10 and R1. Also included in the
XX invention are cDNA sequences encoding the proteins, and antibodies
XX specific for the proteins. The nucleic acid sequences and proteins may be
XX used in the prevention, diagnosis and treatment of diseases associated
XX with inappropriate ADAMTS-N expression. Disorders that may be treated
XX using the nucleic acids, proteins and antibodies include, for example
XX tumour cachexia, inflammation, dermatosparaxis in cattle or Ehlers-Danlos
XX syndrome type VIIC (EDS-VIIC) in humans, erosion of articular cartilage
XX in arthritic (both inflammatory and non-inflammatory) disease,
XX angiogenesis, tumour growth and metastases, and they may also be used for
XX controlling embryogenesis and implantation of fertilised eggs. The
XX present sequence represents human ADAMTS-7. The human ADAMTS-7 gene is
XX located on chromosome 15.
XX SQ Sequence 997 AA;
Query Match 29.7%; Score 212.5; DB 22; Length 997;
Best Local Similarity 37.5%; Pred. No. 3.7e-16;
Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;
OY 11 QEDTRYVLTNLNIGAEILLRDLPSLGAQFRVHLVKMVLTEPEGAPNITANLTSSLLSVCG 70
DB 260 QQVESYVLTINMVGSLPHDPSIGNPHITIVRLVLEDEEDLKITHADNTLKSFK 319
OY 71 WSGTINPEDDTPGHADLVLYITRFDLELPDGNR--QVRGVGTOLGACSPWTSCLITD 128
DB 320 WQKSINMKGDAHPLHDDTAILLTRKDL-CAAMNRPCTLGLSHVAGMCPHRSCLINEDT 378
OY 129 GFDLGVTI 136
DB 379 GLPLAFTV 386
RESULT 11
AAB74944
ID AAB74944 standard; Protein; 1686 AA.
XX
XX AAB74944;
XX
XX 02-JUL-2001 (first entry)
XX Human ADAM type metal protease MDTs1 protein SEQ ID NO:1.
DE
XX Human; a disintegrin and metalloprotease type metal protease; MDTs1;
XX MDTs2; MDTs3; ADAM type metal protease; cytostatic; antiarthritic;
XX cancer; arthritis; arthrosis deformans.
XX
XX Homo sapiens.
XX
XX JP2001008687-A.
XX
XX 16-JAN-2001.
XX
XX 25-JUN-1999; 99JP-0180973.
XX
XX 25-JUN-1999; 99JP-0180973.
XX (YAMA ) YAMANOUCHI PHARM CO LTD.
XX
```

DR WPI; 2001-285362/30.
XX N-PSDB; AAF82149.
XX New metal protease and metal protease gene, for use as a drug for
PT cancers, arthritis and arthrosis deformans
XX Claim 1; Page 12-17; 31pp; Japanese.
XX The present sequence represents a disintegrin and metalloprotease (ADAM)
CC type metal protease designated MDT51, isolated from human. MDT5 proteins
CC have cytostatic and antiarthritic activities. They can be used as a drug
CC for cancers, arthritis and arthrosis deformans.
XX
SQ Sequence 1686 AA;
Query Match 29.7%; Score 212.5; DB 22; Length 1686;
Best Local Similarity 37.5%; Pred. No. 7.9e-16;
Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;
Qy 11 QEDTERVLTNMGAEILLRPSLGAQFRVILVQVILTEGAPNITANLTSSILSYCG 70
Db 260 QPQVESYVLTIMNVAGLFHDSIGNPIHITIVRLVLEDEEDLKITHADNTLKSPCK 319
Qy 71 WSQTINPEDDTPGHADLVLYITRDLPLDGNR--QVRGVTLQGGACSPWMSCLITBDT 128
Db 320 WOKSINMGDAHLPHLDHDTAILTRKDL-CAANNRPCEITGLSHVAGMCPHRSQCSINEDT 378
Qy 129 GPDGLGVTI 136
Db 379 GLPLAFTV 386
RESULT 12
AAE00913
ID AAE00913 standard; Protein; 1686 AA.
XX
AC AAE00913;
XX
DT 04-JUL-2001 (first entry)
XX
DE Human 27875 ADAM-TS protein, alternative version.
XX
KW Human; ADAM-TS; A Disintegrin And Metalloproteinase; antiinflammatory;
KW angio genesis; bronchial asthma; Goodpasture's syndrome; metastasis;
KW heart failure; cardiac hypertrophy; chronic ischaemic heart disease;
KW sickle cell disease nephropathy; urinary tract obstruction; haemostatic;
KW skeletal muscle disorder; myocardial infarction; blood vessel disorder;
KW hypertension; atherosclerosis; vasculitis; renal artery stenosis; mumps;
KW gonorrhea; tuberculosis; syphilis; spermatocytic seminoma; osteoporosis;
KW rhabdomyosarcoma; glomerulonephritis; bone disorder; Paget's disease;
KW rickets; osteomalacia; Hodgkin's disease; gene therapy; antibacterial;
KW candidant; tumour; thymoma; vasotropic; cytostatic; virucide.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= Signal_peptide
FT Modified-site 6..8
FT /note= "Protein kinase C phosphorylation site"
FT Protein 31..1686
FT /label= Human_mature_27875_ADAM-TS_protein
FT Modified-site 55..60
FT /note= "N-myristoylation site"
FT Modified-site 73..75
FT /note= "Protein kinase C phosphorylation site"
FT Domain 78..93
FT /note= "Crystallins beta and gamma Greek key motif"
FT Modified-site 94..97
FT /note= "N-glycosylation site"
FT Modified-site 110..112
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 115..120

FT Modified-site 141..146
FT /note= "N-myristoylation site"
FT Modified-site 147..150
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 159..162
FT /note= "Casein kinase II phosphorylation site"
FT Binding-site 195..197
FT /note= "Cell attachment sequence"
FT Modified-site 214..217
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 214..216
FT /note= "Protein kinase C phosphorylation site"
FT Domain 244..259
FT /label= Metalloproteinase_domain
FT Modified-site 313..315
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 342..345
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 342..344
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 373..376
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 379..384
FT /note= "N-myristoylation site"
FT Domain 385..394
FT /label= Zinc_binding_domain
FT Modified-site 401..404
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 408..411
FT /note= "Amidation site"
FT Modified-site 479..484
FT /note= "N-myristoylation site"
FT Domain 488..567
FT /label= Thrombospondin_domain
FT Modified-site 505..508
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 513..518
FT /note= "N-myristoylation site"
FT Domain 539..545
FT /note= "Growth factor and cytokine metalloproteinase family signature 2 domain"
FT Modified-site 539..544
FT /note= "N-myristoylation site"
FT Domain 541..592
FT /label= Disintegrin_domain
FT Domain 542..592
FT /label= Thrombospondin_domain
FT Modified-site 557..562
FT /note= "N-myristoylation site"
FT Modified-site 569..571
FT /note= "Protein kinase C phosphorylation site"
FT Misc-difference 596
FT /label= Unknown
FT /note= "Encoded by CRC"
FT Modified-site 598..600
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site 605..608
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 614..619
FT /note= "N-myristoylation site"
FT Modified-site 667..672
FT /note= "N-myristoylation site"
FT Binding-site 687..692
FT /note= "Cytochrome C family heme-binding site"
FT Modified-site 688..693
FT /note= "N-myristoylation site"
FT Modified-site 693..696
FT /note= "N-glycosylation site"
FT Modified-site 703..706
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site 716..721
FT /note= "N-myristoylation site"

```

FT Modified-site 740..747 /note= "Tyrosine kinase phosphorylation site"
FT Modified-site 765..770 /note= "N-myristoylation site"
FT Modified-site 774..779 /note= "N-myristoylation site"
FT Modified-site 778..781 /note= "N-glycosylation site"
FT Domain 825..868 /label= Thrombospondin_domain
FT Modified-site 872..875 /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 901..903 /note= "Protein kinase C phosphorylation site"
FT Modified-site 917..920 /note= "Casein kinase II phosphorylation site"
FT Domain 949..994 /label= Thrombospondin_domain
FT Domain 949..988 /label= Thrombospondin_domain
FT Modified-site 950..953 /note= "N-glycosylation site"
FT Modified-site 957..960 /note= "Casein kinase II phosphorylation site"
FT Modified-site 962..964 /note= "Protein kinase C phosphorylation site"
FT Modified-site 971..974 /note= "N-glycosylation site"
FT Modified-site 1005..1010 /note= "N-myristoylation site"
FT Binding-site 1006..1009 /note= "Glycosaminoglycan attachment site"
FT Modified-site 1011..1014 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1035..1037 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1039..1044 /note= "N-myristoylation site"
FT Modified-site 1192..1195 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1263..1252 /note= "N-myristoylation site"
FT Modified-site 1263..1268 /note= "N-myristoylation site"
FT Modified-site 1308..1311 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1358..1363 /note= "N-myristoylation site"
FT Modified-site 1370..1372 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1385..1387 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1397..1400 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1412..1415 /note= "N-glycosylation site"
FT Domain 1415..1463 /label= Thrombospondin_domain
FT Modified-site 1419..1422 /note= "N-glycosylation site"
FT Modified-site 1440..1443 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1440..1442 /note= "Protein kinase C phosphorylation site"
FT Domain 1466..1521 /label= Thrombospondin_domain
FT Modified-site 1470..1473 /note= "N-glycosylation site"
FT Modified-site 1483..1486 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1483..1485 /note= "Protein kinase C phosphorylation site"

```

```

FT Modified-site 1517..1522 /note= "N-myristoylation site"
FT Modified-site 1528..1531 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1528..1530 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1546..1549 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1592..1597 /note= "N-myristoylation site"
FT Modified-site 1599..1601 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1606..1609 /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"

Query Match 29.7%; Score 212.5; DB 22; Length 1686;
Best Local Similarity 37.5%; Pred. No. 7.9e-16;
Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;

QY 11 QEDTERVLTNINIGAEILLRPSLQAFRVHLYKVMVLTEPEGAPNITANLTSSLVSCG 70
Db 260 QPQVESYVLTIMNVAGLFHDPISIGNPIHITIVRLVLEDEEDLKHADNTLKSFK 319
QY 71 WSGTIPEDDTPGHADLVLYTRFDELPDGNR--QVRGVTLGGACSTWKLITEDT 128
Db 320 WQKSINMKGAHPLHHDHTAILLTRKDL-CAAMNRPCTGLSLHVAGMCPHRSCTINEDT 378
QY 129 GFDLGVTI 136
Db 379 GLPLAFTV 386

RESULT 13
AAE00934
ID AAE00934 standard; Protein; 1686 AA.
XX
AC AAE00934;
XX
DT 04-JUL-2001 (first entry)
XX
DE Human 27875 ADAM-TS (a disintegrin and metalloproteinase).
XX
KW Human; ADAM-TS; A Disintegrin And Metalloproteinase; antiinflammatory;
KW angio genesis; bronchial asthma; Goodpasture's syndrome; metastasis;
KW heart failure; cardiac hypertrophy; chronic ischaemic heart disease;
KW sickle cell disease nephropathy; urinary tract obstruction; haemostatic;
KW skeletal muscle disorder; myocardial infarction; blood vessel disorder;
KW hypertension; atherosclerosis; vasculitis; renal artery stenosis; mumps;
KW gonorrhea; tuberculosis; syphilis; spermatocytic seminoma; osteoporosis;
KW rhabdomyosarcoma; glomerulonephritis; bone disorder; Paget's disease;
KW rickets; osteomalacia; Hodgkin's disease; gene therapy; antibacterial;
KW cardiac; tumour; thymoma; vasotropic; cytostatic; virucide.
XX
OS Homo sapiens.
XX
FH Key
FT Peptide 1..30 Location/Qualifiers
FT Modified-site 6..8 /label= Signal_peptide
FT Protein 31..1686 /note= "Protein kinase C phosphorylation site"
FT Modified-site 55..60 /label= Human_mature_27875_ADAM-TS_protein
FT Modified-site 73..75 /note= "N-myristoylation site"
FT Domain 78..93 /note= "Protein kinase C phosphorylation site"
FT Modified-site 94..97 /note= "Crystallins beta and gamma Greek key motif"
FT Modified-site 110..112 /note= "N-glycosylation site"
FT /note= "Protein kinase C phosphorylation site"

```

FT Modified-site 115..120 /note= "N-myristoylation site"
FT Modified-site 141..146 /note= "N-myristoylation site"
FT Modified-site 147..150 /note= "Casein kinase II phosphorylation site"
FT Modified-site 159..162 /note= "Casein kinase II phosphorylation site"
FT Binding-site 195..197 /note= "Cell attachment sequence"
FT Modified-site 214..217 /note= "Casein kinase II phosphorylation site"
FT Modified-site 214..216 /note= "Protein kinase C phosphorylation site"
FT Domain 244..259 /label= Metalloprotease_domain
FT Modified-site 313..315 /note= "Protein kinase C phosphorylation site"
FT Modified-site 342..345 /note= "Casein kinase II phosphorylation site"
FT Modified-site 342..344 /note= "Protein kinase C phosphorylation site"
FT Modified-site 373..376 /note= "Casein kinase II phosphorylation site"
FT Modified-site 379..384 /note= "N-myristoylation site"
FT Domain 385..394 /label= Zinc_binding_domain
FT Modified-site 401..404 /note= "Casein kinase II phosphorylation site"
FT Modified-site 408..411 /note= "Amidation site"
FT Modified-site 479..484 /note= "N-myristoylation site"
FT Domain 488..567 /label= Thrombospondin_domain
FT Modified-site 505..508 /note= "Casein kinase II phosphorylation site"
FT Modified-site 513..518 /note= "N-myristoylation site"
FT Domain 539..545 /note= "Growth factor and cytokine metalloproteinase family signature 2 domain"
FT Modified-site 539..544 /note= "N-myristoylation site"
FT Domain 541..592 /label= Disintegrin_domain
FT Domain 542..592 /label= Thrombospondin_domain
FT Modified-site 557..562 /note= "N-myristoylation site"
FT Modified-site 569..571 /note= "Protein kinase C phosphorylation site"
FT Modified-site 598..600 /note= "Protein kinase C phosphorylation site"
FT Modified-site 605..608 /note= "Casein kinase II phosphorylation site"
FT Modified-site 614..619 /note= "N-myristoylation site"
FT Modified-site 667..672 /note= "N-myristoylation site"
FT Binding-site 687..692 /note= "Cytochrome C family heme-binding site"
FT Modified-site 688..693 /note= "N-myristoylation site"
FT Modified-site 693..696 /note= "N-glycosylation site"
FT Modified-site 703..706 /note= "Casein kinase II phosphorylation site"
FT Modified-site 716..721 /note= "N-myristoylation site"
FT Modified-site 740..747 /note= "Tyrosine kinase phosphorylation site"

FT Modified-site 765..770 /note= "N-myristoylation site"
FT Modified-site 774..779 /note= "N-myristoylation site"
FT Modified-site 778..781 /note= "N-glycosylation site"
FT Domain 825..868 /label= Thrombospondin_domain
FT Modified-site 872..875 /note= "cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 901..903 /note= "Protein kinase C phosphorylation site"
FT Modified-site 917..920 /note= "Casein kinase II phosphorylation site"
FT Domain 949..994 /label= Thrombospondin_domain
FT Domain 949..988 /label= Thrombospondin_domain
FT Modified-site 950..953 /note= "N-glycosylation site"
FT Modified-site 957..960 /note= "Casein kinase II phosphorylation site"
FT Modified-site 962..964 /note= "Protein kinase C phosphorylation site"
FT Modified-site 971..974 /note= "N-glycosylation site"
FT Modified-site 1005..1010 /note= "N-myristoylation site"
FT Binding-site 1006..1009 /note= "Glycosaminoglycan attachment site"
FT Modified-site 1011..1014 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1035..1037 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1039..1044 /note= "N-myristoylation site"
FT Modified-site 1192..1195 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1263..1252 /note= "N-myristoylation site"
FT Modified-site 1263..1268 /note= "N-myristoylation site"
FT Modified-site 1308..1311 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1358..1363 /note= "N-myristoylation site"
FT Modified-site 1370..1372 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1385..1387 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1397..1400 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1412..1415 /note= "N-glycosylation site"
FT Domain 1415..1463 /label= Thrombospondin_domain
FT Modified-site 1419..1422 /note= "N-glycosylation site"
FT Modified-site 1440..1443 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1440..1442 /note= "Protein kinase C phosphorylation site"
FT Domain 1466..1521 /label= Thrombospondin_domain
FT Modified-site 1470..1473 /note= "N-glycosylation site"
FT Modified-site 1483..1486 /note= "Casein kinase II phosphorylation site"
FT Modified-site 1483..1485 /note= "Protein kinase C phosphorylation site"
FT Modified-site 1517..1522 /note= "N-myristoylation site"

```
FT Modified-site 1528..1531
FT /note="Casein kinase II phosphorylation site"
FT Modified-site 1528..1530
FT /note="Protein kinase C phosphorylation site"
FT Modified-site 1546..1549
FT /note="Casein kinase II phosphorylation site"
FT Modified-site 1592..1597
FT /note="N-myristoylation site"
FT Modified-site 1599..1601
FT /note="Protein kinase C phosphorylation site"
FT Modified-site 1606..1609
FT /note="cAMP- and cGMP-dependent protein kinase phosphorylation site"
FT Modified-site 1620..1622
FT /note="Protein kinase C phosphorylation site"
FT Modified-site 1625..1630
FT /note="Protein kinase C phosphorylation site"

Query Match 29.7%; Score 212.5; DB 22; Length 1686;
Best Local Similarity 37.5%; Pred. NO. 7.9e-16;
Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;

QY 11 QEDTERVLTNLNIGALLRDPGLGQAFRVHLVKVILTEPEGAPNITANLTSSLSVCG 70
DB 260 QPOVESYVLTIMNVAAGLFHDPDSIGNPHITIVRLVLEDEEDLKITHADNTLKSFC 319
QY 71 WSGTINPEDDTPGHADLVLYITRFDLELPDGNR--QVRGVLTOLGGACSPWTSCLITE 128
DB 320 WQKSIWKGDAPLHHDTHAILTRKDL-CAAMNRPCTGLSHVAGCQPHRSCSINEDT 378
QY 129 GFDLGVTI 136
DB 379 GLPLAFTV 386

RESULT 14
AAB41379
ID AAB41379 standard; Protein; 1784 AA.
XX
AC AAB41379;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1143 polypeptide sequence SEQ ID NO:2286.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
```

```
XX Shimkets RA, Leach M;
XX WPI; 2000-602362/57.
XX N-PSDB; AAC75588.
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX Claim 11; Page 1678-1682; 5507pp; English.
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antithyroid; and antianaemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX Sequence 1784 AA;
XX
Query Match 29.7%; Score 212.5; DB 21; Length 1784;
Best Local Similarity 37.5%; Pred. NO. 8.6e-16;
Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;

QY 11 QEDTERVLTNLNIGALLRDPGLGQAFRVHLVKVILTEPEGAPNITANLTSSLSVCG 70
DB 263 QPOVESYVLTIMNVAAGLFHDPDSIGNPHITIVRLVLEDEEDLKITHADNTLKSFC 322
QY 71 WSGTINPEDDTPGHADLVLYITRFDLELPDGNR--QVRGVLTOLGGACSPWTSCLITE 128
DB 323 WQKSIWKGDAPLHHDTHAILTRKDL-CAAMNRPCTGLSHVAGCQPHRSCSINEDT 381
QY 129 GFDLGVTI 136
DB 382 GLPLAFTV 389

RESULT 15
AAU72897
ID AAU72897 standard; Protein; 1505 AA.
XX
AC AAU72897;
XX
DT 26-FEB-2002 (first entry)
XX
DE Human metalloprotease partial protein sequence #9.
XX
KW Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;
KW vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;
KW hypotensive; hypotensive; neuroleptic; neuroprotective; anabolic;
KW anorectic; antinflammatory; aspartyl protease; cysteine protease; colon;
KW metalloprotease; serine protease; cancer; haematopoietic; breast;
KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;
KW immune-related disease; cardiovascular disease; neuronal disorder;
KW migraine; sexual dysfunction; mood disorder; attention disorder;
KW cognition disorder; hypotension; hypertension; psychotic disorder;
KW dyskinesia; metabolic disorder; inflammatory disorder.
```



```
XX OS Homo sapiens.
XX PN WO200183782-A2.
XX PD 08-NOV-2001.
XX PF 04-MAY-2001; 2001WO-US14431.
XX PR 04-MAY-2000; 2000US-201879P.
XX PA (SUGE-) SUGEN INC.
XX PI Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
XX PI Payne V;
XX DR WPI: 2002-041502/05.
XX DR N-PSDB; AAS97180.
XX PT Novel protease polypeptide useful for screening for substances that may
XX PT be used to treat, e.g., cancers, immune-related diseases,
XX PT cardiovascular disease, migraine, pain, psychotic and inflammatory
XX PT disorders -
XX PS Claim 28; Figure 2F; 232pp; English.
XX CC The invention relates to an isolated, enriched, or purified protease
XX CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to
XX CC screen for substances (S) that may modulate its activity. Administering
XX CC S (which modulates protease activity in vitro) may be used to treat a
XX CC disease or disorder selected from cancers (e.g., of tissues, of blood or
XX CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,
XX CC brain, ovarian, bladder or kidney), immune-related diseases and
XX CC disorders, cardiovascular disease, brain or neuronal-associated diseases
XX CC (e.g., central or peripheral nervous system diseases, migraine, pain,
XX CC sexual dysfunction, mood disorders, attention disorders, cognition
XX CC disorders, hypotension, hypertension, psychotic disorders, neurological
XX CC disorders and dyskinesias), metabolic disorders and inflammatory
XX CC disorders. (I) may also be useful as a diagnostic tool for a disease or
XX CC disorder such as those above. AAU72876-AAU72910 represent human
XX CC protease amino acid sequences of the invention.
XX SQ Sequence 1505 AA;
Query Match 29.3%; Score 210; DB 23; Length 1505;
Best Local Similarity 34.4%; Pred. No. 1.3e-15;
Matches 45; Conservative 23; Mismatches 63; Indels 0; Gaps 0;
QY 6 VFQAHQEDTERVYLTNLNIGALLRDPGLGQFRVHLVKNWVLTPEGAPNITANLTSSL 65
DB 271 VVSAHGSNLQNYLTLSIVATYIKDPSIGNLIHVVVKLVMIHREEGPVINFDGATTL 330
QY 66 LSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGNRQVRGVTLQGCACSPWTWCLIT 125
DB 331 KNFCSWQQTQNDLDDVHPSHHTAVLITREDICSSKEKCNMLGLSYLGTICDPLQSCFIN 390
QY 126 EDTGFDLGVTI 136
DB 391 EEKGLISAFI 401
Search completed: March 20, 2003, 12:33:30
Job time : 75.7568 secs
```

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:31:32 ; Search time 4.37838 Seconds
(without alignments)
564.722 Million cell updates/sec

Title: SEQID_1

Perfect score: 57

Sequence: 1 AAGGILHLELLV 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues 671580

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	100.0	1427	Q96L37	Q96L37 homo sapien
2	42	73.7	355	Q91WH0	Q91WH0 oryza sativ
3	42	73.7	694	Q9Y4A8	Q9Y4A8 homo sapien
4	39	68.4	175	Q9BP09	Q9BP09 homo sapien
5	38	66.7	151	Q964Q5	Q964Q5 trypanosoma
6	38	66.7	316	Q9UC64	Q9UC64 homo sapien
7	38	66.7	339	Q96FK1	Q96FK1 sulfolobus
8	38	66.7	384	Q96CF5	Q96CF5 homo sapien
9	38	66.7	489	Q60416	Q60416 homo sapien
10	38	66.7	490	Q96FV4	Q96FV4 homo sapien
11	38	66.7	1081	Q68831	Q68831 bacteroides
12	37	64.9	116	Q9YAE8	Q9YAE8 aeropyrum p
13	37	64.9	322	Q93F08	Q93F06 prochloroco
14	37	64.9	322	Q93P05	Q93P05 prochloroco
15	37	64.9	322	Q93P04	Q93P04 prochloroco
16	37	64.9	352	Q07295	Q07295 prochloroco

17	37	64.9	356	17	Q97WN3	Q97WN3 sulfolobus
18	37	64.9	388	16	Q8XVW1	Q8XVW1 ralsstonia s
19	37	64.9	454	2	Q93P65	Q93P65 pseudomonas
20	37	64.9	477	13	Q8QHI9	Q8QHI9 oncorhynchus
21	37	64.9	555	10	Q9FI58	Q9FI58 arabidopsis
22	37	64.9	555	10	Q94BN5	Q94BN5 arabidopsis
23	37	64.9	576	2	Q8TR5	Q8TR5 uncultured
24	37	64.9	732	4	Q96T17	Q96T17 homo sapien
25	37	64.9	876	16	Q93IG6	Q93IG6 mycobacteri
26	37	64.9	1326	5	Q9W4N4	Q9W4N4 drosophila
27	37	64.9	1423	5	Q95YP5	Q95YP5 leishmania
28	36	63.2	233	16	Q86804	Q86804 streptomyc
29	36	63.2	335	12	Q8V6J1	Q8V6J1 trocara vir
30	36	63.2	335	12	Q8V6J0	Q8V6J0 trocara vir
31	36	63.2	451	5	Q95T70	Q95T70 drosophila
32	36	63.2	470	5	Q9VG87	Q9VG87 drosophila
33	36	63.2	480	4	Q96DD2	Q96DD2 homo sapien
34	36	63.2	505	16	Q8UAA5	Q8UAA5 agrobacteri
35	36	63.2	597	11	Q9NVR2	Q9NVR2 homo sapien
36	36	63.2	597	11	Q9DSU1	Q9DSU1 mus musculu
37	36	63.2	842	3	Q9P6K4	Q9P6K4 schizosacch
38	36	63.2	1142	10	Q40263	Q40263 mesotaenium
39	35	61.4	151	5	O18573	O18573 andrena eri
40	35	61.4	165	17	Q981B4	Q981B4 sulfolobus
41	35	61.4	170	10	Q94KG6	Q94KG6 phaseolus v
42	35	61.4	170	10	Q94KG3	Q94KG3 phaseolus v
43	35	61.4	200	2	Q8RPP4	Q8RPP4 legionella
44	35	61.4	218	16	Q9RDP9	Q9RDP9 streptomyc
45	35	61.4	231	10	Q94KP4	Q94KP4 phaseolus v

ALIGNMENTS

RESULT 1
Q96L37 PRELIMINARY; PRT; 1427 AA.
ID Q96L37
AC Q96L37; (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Von Willebrand factor-cleaving protease precursor.
GN ADAMTS13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX PubMed=11557746;
RA Zheng X., Chung D., Takayama T.K., Majerus E.M., Sadler J.E.,
RA Fujikawa K.;
RT "Structure of von Willebrand Factor-cleaving Protease (ADAMTS13), a
RT Metalloprotease Involved in Thrombotic Thrombocytopenic Purpura.";
RL J. Biol. Chem. 276:41059-41063 (2001).
DR EMBL; AY055376; AAL17652.1;
DR MEROPS; M12.241;
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSF1.
DR InterPro; IPR000130; Zn.Mtpeptdse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tse1; 4.
DR PROSITE; PS50215; ADAM_MERO; 1.
DR PROSITE; PS50092; TSF1; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 75 1427 VON WILLEBRAND FACTOR-CLEAVING PROTEASE.
SQ SEQUENCE 1427 AA; 153632 MW; EB1BC3AABCA4442 CRC64;

Query Match 100.0%; Score 57; DB 4; Length 1427;
Best Local Similarity 100.0%; Pred. No. 0.04;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGILHLELLV 12

Db 75 AAGGILHLELLV 86

RESULT 2

Q9LWH0 PRELIMINARY; PRT; 355 AA.

AC Q9LWH0;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE Hypothetical protein.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzae; Oryza.

OX NCBI_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. NIPPONBARE;

RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC

RI Clone: P0675A05."

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AP002071; BAA95889.1; -.

KW Hypothetical protein.

SQ SEQUENCE 355 AA; 37762 MW; C5AD33502A9D3732 CRC64;

Query Match 73.7%; Score 42; DB 10; Length 355;

Best Local Similarity 66.7%; Pred. No. 7.4; Mismatches 2; Indels 0; Gaps 0;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELLV 12

Db 49 ASGGVHLVLFV 60

RESULT 3

Q9Y4A8 PRELIMINARY; PRT; 694 AA.

AC Q9Y4A8; Q9UQA4;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE NF-E2-related factor 3.

GN NRF3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Stoesz S.P., Liu S., Pickett C.B.;

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Blank V., Andrews N.C.;

RT "NRF3, a basic leucine zipper protein interacting with MAPG.";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 295-694 FROM N.A.

RC TISSUE=BONE MARROW, AND PLACENTA;

RX MEDLINE=99156926; PubMed=10037736;

RA Kobayashi A., Ito E., Toki T., Kogame K., Takahashi S., Igarashi K.;

RA Hayashi N., Yamamoto M.;

RT "Molecular cloning and functional characterization of a new cap'n'

RT collar family transcription factor Nrf3.";

RL J. Biol. Chem. 274:6443-6452 (1999).

CC -/- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -/- SIMILARITY: BELONGS TO THE BZIP FAMILY.

EMBL; AF135116; AAF61415.1; -.

EMBL; AF134891; AAF61404.1; -.

EMBL; AF133059; AAG43275.1; -.

EMBL; AB010812; BAA76288.1; -.

DR HSSP; P34707; ISKN.

DR InterPro; IPR004827; TF_bZIP.

DR Pfam; PF00170; bZIP; 1.

DR SMART; SM00338; BRLZ; 1.

DR PROSITE; PS00036; BZIP_BASIC; 1.

KW DNA-binding; Nuclear protein.

SQ SEQUENCE 694 AA; 76153 MW; DAD6D883FCCBA982 CRC64;

Query Match 73.7%; Score 42; DB 4; Length 694;

Best Local Similarity 66.7%; Pred. No. 15; Mismatches 2; Indels 0; Gaps 0;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELLV 12

Db 10 AGGGLHLTL 21

RESULT 4

Q9BP09 PRELIMINARY; PRT; 175 AA.

AC Q9BP09;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE Unknown (protein for MGC:4093) (protein for MGC:2435).

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=LUNG;

RA Strausberg R.;

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC004444; AAH04444.1; -.

DR EMBL; BC004157; AAH04157.1; -.

SQ SEQUENCE 175 AA; 19279 MW; FCECD4E3207E9765 CRC64;

Query Match 68.4%; Score 39; DB 4; Length 175;

Best Local Similarity 70.0%; Pred. No. 13; Mismatches 2; Indels 1; Gaps 0;

Matches 7; Conservative 2; Mismatches 1; Indels 1; Gaps 0;

Qy 1 AAGGILHLEL 10

Db 152 AAGGTVHLEI 161

RESULT 5

Q964Q5 PRELIMINARY; PRT; 151 AA.

AC Q964Q5;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Putative ubiquitin-conjugating enzyme.

OS Trypanosoma cruzi.

OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

OX NCBI_TaxID=5693;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CL-BRENNER;

RA Pravia C.A., Bonfempi E.;

RA "A novel ubiquitin-conjugating enzyme of trypanosoma cruzi.";

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

CC -/- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER

CC PROTEINS (BY SIMILARITY).

CC -/- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE = AMP +

CC -/- DIPHOSPHATE + PROTEIN N-UBIQUITYLLYSINE.

CC -/- PATHWAY: SECOND STEP IN UBIQUITIN CONJUGATION.

RN SEQUENCE FROM N.A.
 RA Rohlfing T., Wohldmann P., Antoniou B., Bauer C., O'Neal D.;
 RL "The sequence of Homo sapiens BAC clone CTA-27603."
 RT Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC004668; AAC15866.1; -.
 DR MEROPS; M16.003; -.
 DR InterPro; IPR001431; Peptidase M16.
 DR Pfam; PF00675; Peptidase_M16; 1.
 DR PROSITE; PS00143; INSULINASE; 1.
 SQ SEQUENCE 489 AA; 54366 MW; 79250D016E60CFEE CRC64;

Query Match 66.7%; Score 38; DB 4; Length 489;
 Best Local Similarity 72.7%; Pred. No. 62;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
 ||||: ||||
 Db 243 AAGGVSHDELL 253

RESULT 10

Q96FV4
 ID Q96FV4 PRELIMINARY; PRT; 490 AA.
 AC Q96FV4;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Unknown (protein for MGC:13691).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A.
 RA Tissue=SKETAL MUSCLE;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC010398; AAH10398.1; -.
 DR MEROPS; M16.975; -.
 DR InterPro; IPR001431; Peptidase M16.
 DR Pfam; PF00675; Peptidase M16; 1.
 DR PROSITE; PS00143; INSULINASE; UNKNOWN 1.
 SQ SEQUENCE 490 AA; 54191 MW; 8999F3F57A4D14E1 CRC64;

Query Match 66.7%; Score 38; DB 4; Length 490;
 Best Local Similarity 72.7%; Pred. No. 62;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
 ||||: ||||
 Db 243 AAGGVSHDELL 253

RESULT 11

O68831
 ID O68831 PRELIMINARY; PRT; 1081 AA.
 AC O68831;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Surface antigen BspA.
 GN BSPA.
 OS Bacteroides forsythus.
 OC Bacteria; CF8 group; Bacteroidetes; Bacteroidales; Bacteroidaceae;
 OC Bacteroides.
 OX NCBI_TaxID=28112;
 RN [1]
 SQ SEQUENCE FROM N.A.

RC STRAIN=ATCC43037;
 RX MEDLINE=99043895; PubMed=9826345;
 RA Sharma A., Sojar H.T., Glurich I., Honma K., Kuramitsu H.K.,
 RA Genco R.J.;
 RT "Cloning, expression, and sequencing of a cell surface antigen
 RT containing a leucine-rich repeat motif from Bacteroides forsythus ATCC
 RT 43037."
 RL Infect. Immun. 66:5703-5710(1998).
 DR EMBL; AF054892; AAC89625.1; -.
 DR InterPro; IPR003343; B19_2.
 DR Pfam; PF02368; B19_2; 1.
 DR PRINTS; PR01611; LIMP11.
 SQ SEQUENCE 1081 AA; 113908 MW; 6E5E36FA0ACDBDSC CRC64;

Query Match 66.7%; Score 38; DB 2; Length 1081;
 Best Local Similarity 80.0%; Pred. No. 1.4e+02;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLEL 10
 ||||: ||||
 Db 1023 AAGGALHLL 1032

RESULT 12

Q9YAE8
 ID Q9YAE8 PRELIMINARY; PRT; 116 AA.
 AC Q9YAE8;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein APE1991.
 GN APE1991.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
 OC Desulfurococaceae; Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KJ;
 RX MEDLINE=9910339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaï A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kuehida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1."
 RL DNA Res. 6:83-101(1999).
 DR EMBL; AP000063; BAA81001.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 116 AA; 12506 MW; 2ADE1A5B2B82FA8A CRC64;

Query Match 64.9%; Score 37; DB 17; Length 116;
 Best Local Similarity 77.8%; Pred. No. 21;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGGILHLE 9
 ||||: ||||
 Db 44 AAGGLQHLE 52

RESULT 13

Q93P06
 ID Q93P06 PRELIMINARY; PRT; 322 AA.
 AC Q93P06;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Chlorophyll a/b-binding light-harvesting protein Pcb
 DE (Fragment).
 GN PCB.
 OS Prochlorococcus marinus.

OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
 OC Prochlorococcus.
 OX NCBI_TaxID=1219;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=SB;
 RA Garczarek L., van der Staay G.W.M., Hess W.R., Le Gall F.,
 RA Partensky F.;
 RA "Expression and phylogeny of the multiple antenna genes of the low-
 RT light adapted strain Prochlorococcus marinus SS120
 RT (Oxyphotobacteria).";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF354645; AAK69279.1; -.
 DR InterPro; IPR000932; PSIIProt.
 DR InterPro; IPR000408; Reg_chr_condens.
 DR Pfam; PF00421; PSII; 1.
 DR PROSITE; PS00626; RCC1_2; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 322
 FT NON_TER 322
 SQ SEQUENCE 322 AA; 35495 MW; B0762A1DD6279032 CRC64;

Query Match 64.9%; Score 37; DB 2; Length 322;
 Best Local Similarity 72.7%; Pred. No. 62;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAGGILHLELL 11
 ||||:||||
 Db 98 AAGGLLHSLLL 108

RESULT 14

ID Q93P05 PRELIMINARY; PRT; 322 AA.
 AC Q93P05;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Chlorophyll a/b-binding light-harvesting protein Pcb
 DE (Fragment).
 GN PCB.
 OS Prochlorococcus marinus.
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
 OC Prochlorococcus.
 OX NCBI_TaxID=1219;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=GP2;
 RA Garczarek L., van der Staay G.W.M., Hess W.R., Le Gall F.,
 RA Partensky F.;
 RA "Expression and phylogeny of the multiple antenna genes of the low-
 RT light adapted strain Prochlorococcus marinus SS120
 RT (Oxyphotobacteria).";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF354646; AAK69280.1; -.
 DR InterPro; IPR000932; PSIIProt.
 DR InterPro; IPR000408; Reg_chr_condens.
 DR Pfam; PF00421; PSII; 1.
 DR PROSITE; PS00626; RCC1_2; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 322
 FT NON_TER 322
 SQ SEQUENCE 322 AA; 35395 MW; DF16855EDB565ADE CRC64;

Query Match 64.9%; Score 37; DB 2; Length 322;
 Best Local Similarity 72.7%; Pred. No. 62;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAGGILHLELL 11
 ||||:||||
 Db 98 AAGGLLHSLLL 108

RESULT 15

Q93P04

ID Q93P04 PRELIMINARY; PRT; 322 AA.
 AC Q93P04;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Chlorophyll a/b-binding light-harvesting protein Pcb
 DE (Fragment).
 GN PCB.
 OS Prochlorococcus marinus.
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
 OC Prochlorococcus.
 OX NCBI_TaxID=1219;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=TAK9803-2;
 RA Garczarek L., van der Staay G.W.M., Hess W.R., Le Gall F.,
 RA Partensky F.;
 RA "Expression and phylogeny of the multiple antenna genes of the low-
 RT light adapted strain Prochlorococcus marinus SS120
 RT (Oxyphotobacteria).";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF354647; AAK69281.1; -.
 DR InterPro; IPR000932; PSIIProt.
 DR InterPro; IPR000408; Reg_chr_condens.
 DR Pfam; PF00421; PSII; 1.
 DR PROSITE; PS00626; RCC1_2; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 322
 FT NON_TER 322
 SQ SEQUENCE 322 AA; 35386 MW; BB336C0143ED3DCF CRC64;

Query Match 64.9%; Score 37; DB 2; Length 322;
 Best Local Similarity 72.7%; Pred. No. 62;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAGGILHLELL 11
 ||||:||||
 Db 98 AAGGLLHSLLL 108

Search completed: March 20, 2003, 12:34:28
 Job time : 7.37838 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:31:05 ; Search time 6.24324 Seconds
(without alignments)
256.118 Million cell updates/sec

Title: SEQID_1

Perfect score: 57

Sequence: 1 AAGGILHLELLV 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	100.0	12	AAE24446	Human Von Willebra
2	57	100.0	15	AAE24447	Human Von Willebra
3	57	100.0	15	AAE24448	Human Von Willebra
4	57	100.0	148	AAE24449	Human Von Willebra
5	57	100.0	203	AAE24450	Amino acid sequenc
6	57	100.0	242	AAU97641	Human aggrecanase
7	57	100.0	933	AAU79217	Human ADAM-TS-like
8	57	100.0	1120	AAU63829	Amino acid sequenc
9	57	100.0	1353	AAE24449	Human Von Willebra
10	57	100.0	1416	ABB04153	Human ADAMTS-M pol

11	57	100.0	1427	23	AAE24450	Human Von Willebra
12	45	78.9	9	23	AAE24453	Human Von Willebra
13	42	73.7	669	22	ABG11310	Novel human diagno
14	42	73.7	694	21	AAB42910	Human ORFX ORF2674
15	42	73.7	694	22	ABG11309	Novel human diagno
16	42	73.7	697	22	ABG15217	Novel human diagno
17	39	68.4	175	22	AAE24459	Human polypeptide
18	39	68.4	212	22	AAW41245	Human polypeptide
19	38	66.7	460	21	AAE28098	Human secreted pro
20	38	66.7	462	20	AAE28097	Mitochondrial proc
21	38	66.7	489	20	AAW88114	Human secreted pro
22	38	66.7	489	21	AAW44658	Human mitochondria
23	38	66.7	489	21	AAW57558	Human protein sequ
24	38	66.7	489	22	AAE24545	Human protein sequ
25	37	64.9	112	21	AAE23793	Arabidopsis thalia
26	37	64.9	124	21	AAE23792	Arabidopsis thalia
27	37	64.9	139	22	ABG28850	Novel human diagno
28	37	64.9	154	21	AAE23791	Arabidopsis thalia
29	37	64.9	204	22	ABG28814	Novel human diagno
30	37	64.9	376	21	AAE24460	Arabidopsis thalia
31	37	64.9	452	21	AAE24459	Arabidopsis thalia
32	37	64.9	555	21	AAE24458	Arabidopsis thalia
33	37	64.9	732	22	AAE24457	Human protein sequ
34	37	64.9	819	22	ABG24680	Novel human diagno
35	37	64.9	819	22	ABG24678	Novel human diagno
36	37	64.9	876	20	AAE24456	Mycobacterium spec
37	37	64.9	886	20	AAE24455	Mycobacterium spec
38	37	64.9	1326	22	ABE61156	Drosophila melanog
39	36	63.2	7	23	AAE24454	Human Von Willebra
40	36	63.2	79	22	AAU44261	Propionibacterium
41	36	63.2	106	22	ABG00377	Novel human diagno
42	36	63.2	115	21	AAE24453	Human secreted pro
43	36	63.2	470	22	ABE71838	Drosophila melanog
44	36	63.2	597	22	AAE24452	Human protein sequ
45	36	63.2	665	22	AAE24451	Human colon cancer

ALIGNMENTS

RESULT 1
AAE24446
ID AAE24446 standard; peptide; 12 AA.
XX
AC AAE24446;
XX
DT 04-OCT-2002 (first entry)
XX
DE Human Von Willebrand factor-cleaving protease (VWF-cp) peptide #1.
XX
KW Human; Von Willebrand factor-cleaving protease; VWF-cp; therapy; enzyme;
KW transgenic animal; immunisation; thromboembolic disease; pre-eclampsia;
KW thrombotic thrombocytopenic purpura; TTP; Henoch-Schönlein purpura;
KW thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
KW transgenic; anticoagulant.
XX
OS Homo sapiens.
PN OS
PN WO200242441-A2.
XX
PD 30-MAY-2002.
XX
PF 20-NOV-2001; 2001WO-EP13391.
XX
PR 22-NOV-2000; 2000US-0721254.
PR 12-APR-2001; 2001US-0833328.
XX
PA (BAXT) BAXTER AG.
XX
PI Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
PI Scheiflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
PI Zimmermann K, Voelkel D;
XX

DR WPI; 2002-479950/51.

XX Novel isolated or substantially purified Von Willebrand factor-cleaving

PT protease, useful for producing preparation for therapy of thrombosis

PT and thromboembolic disease such as thrombotic thrombocytic purpura

XX

XX Claim 1; Page 38; 93pp; English.

XX The invention relates to an isolated or substantially pure Von Willebrand

CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for

CC purifying vWF which involves providing vWF-cp as a ligand, contacting a

CC solution comprising vWF with the polypeptide ligand under conditions

CC where vWF is bound to the ligand and recovering from the ligand purified

CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies

CC which involves immunising an animal with vWF-cp and isolating the anti-

CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for

CC producing a preparation of prophylaxis and therapy of thrombosis and

CC thromboembolic disease such as thrombotic thrombocytic purpura (TTP),

CC Henoch-Schonlein purpura, preeclampsia, neonatal thrombocytopenia or

CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing

CC plasmatic or recombinantly produced vWF. The invention is useful for

CC construction expression systems and generating transgenic animals which

CC express the polypeptide in vivo. The present sequence is human vWF-cp

XX peptide.

XX

SQ Sequence 12 AA;

Query Match 100.0%; Score 57; DB 23; Length 12;

Best Local Similarity 100.0%; Pred. NO. 0.00083;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGILHLELLV 12

Db 1 AAGGILHLELLV 12

RESULT 2

AAE24447

ID AAE24447 standard; peptide; 15 AA.

AC AAE24447;

XX

XX 04-OCT-2002 (first entry)

DT Human Von Willebrand factor-cleaving protease N-terminal peptide #1.

DE

XX Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;

XX transgenic animal; immunisation; thromboembolic disease; preeclampsia;

XX thrombotic thrombocytic purpura; TTP; Henoch-Schonlein purpura;

XX thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;

XX transgenic; anticoagulant.

OS

XX Homo sapiens.

XX

XX WO200242441-A2.

PN

XX 30-MAY-2002.

PD

XX

XX 20-NOV-2001; 2001WO-EPI3391.

PF

XX

XX 22-NOV-2000; 2000US-0721254.

PR

XX 12-APR-2001; 2001US-0833328.

PR

XX (BAXT) BAXTER AG.

PA

XX Laemmle B. Gerritsen HE, Furlan M, Turecek P, Schwarz H;

XX Scheifflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;

XX Zimmermann K, Voelkel D;

XX WPI; 2002-479950/51.

XX

XX N-PSDB; AAD39330.

XX

XX Novel isolated or substantially purified Von Willebrand factor-cleaving

PT protease, useful for producing preparation for therapy of thrombosis

PT and thromboembolic disease such as thrombotic thrombocytic purpura

PS

PT protease, useful for producing preparation for therapy of thrombosis

PT and thromboembolic disease such as thrombotic thrombocytic purpura

PS

XX Claim 1; Page 32; 93pp; English.

XX The invention relates to an isolated or substantially pure Von Willebrand

CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for

CC purifying vWF which involves providing vWF-cp as a ligand, contacting a

CC solution comprising vWF with the polypeptide ligand under conditions

CC where vWF is bound to the ligand and recovering from the ligand purified

CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies

CC which involves immunising an animal with vWF-cp and isolating the anti-

CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for

CC producing a preparation of prophylaxis and therapy of thrombosis and

CC thromboembolic disease such as thrombotic thrombocytic purpura (TTP),

CC Henoch-Schonlein purpura, preeclampsia, neonatal thrombocytopenia or

CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing

CC plasmatic or recombinantly produced vWF. The invention is useful for

CC construction expression systems and generating transgenic animals which

CC express the polypeptide in vivo. The present sequence is human vWF-cp

XX peptide.

XX

SQ Sequence 15 AA;

Query Match 100.0%; Score 57; DB 23; Length 15;

Best Local Similarity 100.0%; Pred. NO. 0.0011;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGILHLELLV 12

Db 1 AAGGILHLELLV 12

RESULT 3

AAE24454

ID AAE24454 standard; peptide; 15 AA.

AC AAE24454;

XX

XX 04-OCT-2002 (first entry)

DT Human Von Willebrand factor-cleaving protease N-terminal peptide #5.

DE

XX Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;

XX transgenic animal; immunisation; thromboembolic disease; preeclampsia;

XX thrombotic thrombocytic purpura; TTP; Henoch-Schonlein purpura;

XX thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;

XX transgenic; anticoagulant.

OS

XX Homo sapiens.

XX

XX WO200242441-A2.

PN

XX 30-MAY-2002.

PD

XX

XX 20-NOV-2001; 2001WO-EPI3391.

PF

XX

XX 22-NOV-2000; 2000US-0721254.

PR

XX 12-APR-2001; 2001US-0833328.

PR

XX (BAXT) BAXTER AG.

PA

XX Laemmle B. Gerritsen HE, Furlan M, Turecek P, Schwarz H;

XX Scheifflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;

XX Zimmermann K, Voelkel D;

XX WPI; 2002-479950/51.

XX

XX Novel isolated or substantially purified Von Willebrand factor-cleaving

PT protease, useful for producing preparation for therapy of thrombosis

PT and thromboembolic disease such as thrombotic thrombocytic purpura

PS

XX Example 2; Page 30; 93pp; English.

XX The invention relates to an isolated or substantially pure Von Willebrand
CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
CC solution comprising vWF with the polypeptide ligand under conditions
CC where vWF is bound to the ligand and recovering from the ligand purified
CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
CC which involves immunising an animal with vWF-cp and isolating the anti-
CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
CC producing a preparation of prophylaxis and therapy of thrombosis and
CC thromboembolic disease such as thrombotic thrombocytopenia or
CC Henoch-Schönlein purpura, pre-eclampsia, neonatal thrombocytopenia or
CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
CC plasmatic or recombinantly produced vWF. The invention is useful for
CC construction expression systems and generating transgenic animals which
CC express the polypeptide in vivo. The present sequence is human vWF-cp
CC N-terminal peptide.

XX SQ Sequence 15 AA;
Query Match 100.0%; Score 57; DB 23; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLV 12
Db 1 AAGGILHLELLV 12
|||||

RESULT 4
AAE24478
ID AAE24448 standard; Protein; 148 AA.
XX AAE24448;
AC AAE24448;
XX 04-OCT-2002 (first entry)
DT Human Von Willebrand factor-cleaving protease fragment #1.
DE Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;
XX transgenic animal; immunisation; thromboembolic disease; pre-eclampsia;
KW thrombotic thrombocytopenic purpura; TTP; Henoch-Schönlein purpura;
KW thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
KW transgenic; anticoagulant.
XX Homo sapiens.
OS WO200242441-A2.
XX 30-MAY-2002.
XX 20-NOV-2001; 2001WO-EPI3391.
XX 22-NOV-2000; 2000US-0721254.
XX 12-APR-2001; 2001US-0833328.
XX (BAXT) BAXTER AG.
XX Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
PI Scheiflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
PI Zimmermann K, Voelkel D;
XX WPI; 2002-479950/51.
XX N-PSDB; AAD39331.
XX Novel isolated or substantially purified Von Willebrand factor-cleaving
XX protease, useful for producing preparation for therapy of thrombosis
XX and thromboembolic disease such as thrombotic thrombocytopenic purpura
XX Claim 1; Fig 3; 93pp; English.
PS The invention relates to an isolated or substantially pure Von Willebrand
XX factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
CC

CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
CC solution comprising vWF with the polypeptide ligand under conditions
CC where vWF is bound to the ligand and recovering from the ligand purified
CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
CC which involves immunising an animal with vWF-cp and isolating the anti-
CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
CC producing a preparation of prophylaxis and therapy of thrombosis and
CC thromboembolic disease such as thrombotic thrombocytopenic purpura (TTP),
CC Henoch-Schönlein purpura, pre-eclampsia, neonatal thrombocytopenia or
CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
CC plasmatic or recombinantly produced vWF. The invention is useful for
CC construction expression systems and generating transgenic animals which
CC express the polypeptide in vivo. The present sequence is human vWF-cp
CC fragment.

XX SQ Sequence 148 AA;
Query Match 100.0%; Score 57; DB 23; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLV 12
Db 1 AAGGILHLELLV 12
|||||

RESULT 5
AAG63826
ID AAG63826 standard; Protein; 203 AA.
XX AAG63826;
AC AAG63826;
XX 29-OCT-2001 (first entry)
DT Amino acid sequence of a human zdtnt5 polypeptide.
DE Human; zdtnt5; anti-angiogenic; intestinal polypeptide; wound healing;
XX extracellular matrix interaction; tumour suppression; gamete maturation;
KW immunologic recognition; gastrointestinal irradiation; chemotherapy;
KW proteolysis; apoptosis; angiogenesis; infection; cell adhesion;
KW cell fusion; cell signalling; tumour; Crohn's disease; melanoma; trauma;
KW inflammatory bowel disease; food poisoning; degenerative disease;
KW inflammation; fertility; gamete maturation; epithelial disorder.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH Modified-site 69
FT Modified-site /note= "potential N-linked glycosylation site"
FT Modified-site 73
FT Modified-site /note= "potential N-linked glycosylation site"
FT Region 151..161
FT /note= "zinc-binding motif"
XX WO200159112-A1.
XX 16-AUG-2001.
XX 09-FEB-2001; 2001WO-US04198.
XX 10-FEB-2000; 2000US-0501806.
XX (ZYMO) ZYMOGENETICS INC.
XX Holloway JL, Sheppard PO;
PI WPI; 2001-522477/57.
XX N-PSDB; AAH74759.
XX New anti-angiogenic intestinal polypeptides; zdtnt5 polypeptides, which
XX are members of disintegrin proteases, for modulating extracellular
XX matrix interaction, tumour suppression and wound healing

PS Claim 1; Page 73-74; 92pp; English.

XX The present sequence represents a human zdn5 polypeptide. The zdn5 polypeptide is an anti-angiogenic intestinal polypeptide. zdn5 is used for modulating extracellular matrix interactions. zdn5 polypeptide is useful as a tool for identifying new family members of polypeptides. zdn5 polynucleotides are useful as probes or primers to clone 5' non-coding regions of zdn5 gene. zdn5 polypeptides are used for tumour suppression, gamete maturation, immunologic recognition, and growth and differentiation either working in isolation or in conjunction with other molecules in colon, small intestine, fetal lung, testis and B-cells. zdn5 polypeptides are also useful for promoting wound healing. In the treatment of disorders associated with recovery after gastrointestinal irradiation, chemotherapy or antibody use, as anti-infectives, and extracellular matrix repair and remodeling. The polypeptides are also useful for modulating proteolysis, apoptosis, angiogenesis, infection, cell adhesion, cell fusion and signalling. The polypeptides are also useful for treating tumour formation, Crohn's disease, inflammatory bowel disease, food poisoning, melanoma, degenerative diseases, disorders related to immunity, inflammation, fertility, gamete maturation, immunology, trauma and epithelial disorders.

XX SQ Sequence 203 AA;

Query Match 100.0%; Score 57; DB 22; Length 203;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGILHLELLV 12
Db 2 AAGGILHLELLV 13
|||||

RESULT 6
AAU97641
ID AAU97641 standard; Protein; 242 AA.
AC AAU97641;
XX 27-AUG-2002 (first entry)
XX Human aggrecanase protein.
XX Aggrecanase; osteopathetic; antiarthritic; antiinflammatory; human; cartilage; osteoarthritis; inflammatory disease; enzyme.
XX Homo sapiens.
XX WO200233093-A2.
XX 25-APR-2002.
XX 17-OCT-2001; 2001WO-US32458.
XX 18-OCT-2000; 2000US-241469P.
XX (GEMY) GENETICS INST INC.
XX Racie LA, Twine NC, Agostino MJ, Wolfman NM, Morris EA;
PI WPI; 2002-452389/48.
DR N-PSDB; ABK52579.
XX Novel purified aggrecanase polypeptide useful for developing inhibitors and antibodies to the aggrecanase polypeptide, which are useful for treating aggrecanase-associated condition such as osteoarthritis -
XX Claim 1; Page 33-34; 41pp; English.
XX This invention relates to the cDNA and protein sequences of a novel human aggrecanase polypeptide. The protein of the invention may be used to inhibit the proteolytic activity of aggrecanase, or to inhibit

CC the aggrecanase-mediated cleavage of aggrecan in cartilage. The protein of the invention is useful for developing inhibitors of aggrecanase protein. The cDNA sequence encoding the aggrecanase protein of the invention is useful for designing probes for obtaining DNA sequences encoding other aggrecanase molecules. The cDNA sequence is also useful for detecting mRNA encoding aggrecanase in a given cell population, and thus for detecting or diagnosing genetic disorders involving the aggrecanase, or disorders involving cellular, organ or tissue disorders in which aggrecanase is irregularly transcribed or expressed. The DNA sequences may also be useful for preparing vectors for gene therapy applications. An inhibitor of the protein is useful in treating conditions characterised by degradation of articular cartilage, by blocking the enzyme's proteolytic activity. An aggrecanase protein inhibitor and a method for inhibition of its activity are useful for treating various aggrecanase-associated conditions including osteoarthritis and other inflammatory diseases. The present sequence represents the human aggrecanase protein of the invention.

XX SQ Sequence 242 AA;

Query Match 100.0%; Score 57; DB 23; Length 242;
Best Local Similarity 100.0%; Pred. No. 0.023;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGILHLELLV 12
Db 42 AAGGILHLELLV 53
|||||

RESULT 7
AAU79217
ID AAU79217 standard; Protein; 933 AA.
AC AAU79217;
XX 15-JUL-2002 (first entry)
XX Human ADAM-TS-like protein.
XX Human; ADAM-TS-like protein; cardiovascular disorder; angina; vascular system; congestive heart failure; myocardial infarction; ischaemic heart disease; arrhythmia; hypertensive vascular disease; secondary arterial hypertension; peripheral vascular disease; embolism; chronic peripheral arterial occlusive disease; acute arterial thrombosis; inflammatory vascular disorder; chronic obstructive pulmonary disease; liver disorder.
XX Homo sapiens.
XX WO200226999-A2.
XX 04-APR-2002.
XX 26-SEP-2001; 2001WO-EP11124.
XX 28-SEP-2000; 2000US-235881P.
XX 25-JUL-2001; 2001US-307393P.
XX (FARB) BAYER AG.
XX Xiao Y;
XX WPI; 2002-383274/41.
DR N-PSDB; ABK49356.
XX New purified human ADAM-TS-like protein, useful for identifying modulators of protein activity for treating cardiovascular or liver disorder or chronic obstructive pulmonary disease -
XX Claim 25; Fig 2; 106pp; English.
XX The invention relates to a human ADAM-TS-like protein and the

CC polynucleotide encoding it. The protein of the invention is useful for
 CC treating cardiovascular disorders including diseases of the heart and
 CC vascular system, such as congestive heart failure, myocardial infarction,
 CC ischaemic heart diseases (e.g., stable angina, unstable angina), atrial
 CC and ventricular arrhythmia, hypertensive vascular diseases (e.g., all
 CC kinds of secondary arterial hypertension), and peripheral vascular
 CC diseases (e.g., chronic peripheral arterial occlusive disease, acute
 CC arterial thrombosis and embolism, inflammatory vascular disorders, etc),
 CC liver disorders and chronic obstructive pulmonary disease. The sequences
 CC are useful in diagnostic assays for detecting diseases and abnormalities
 CC or susceptibility to diseases and abnormalities related to the presence
 CC of mutations in the nucleic acid sequences which encode the protein. The
 CC sequences are also useful for modulating ADAM-TS-like protein activity in
 CC a disease condition. This sequence represents the human ADAM-TS-like
 CC protein.

XX Sequence 933 AA;

Query Match 100.0%; Score 57; DB 23; Length 933;
 Best Local Similarity 100.0%; Pred. No. 0.1;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILHLELLV 12
 DB 78 AAGGILHLELLV 89
 |||||

RESULT 8

AAG63829
 ID AAG63829 standard; Protein; 1120 AA.

XX AAG63829;

XX 29-OCT-2001 (first entry)

DE Amino acid sequence of a human zdtint5 polypeptide.

XX Human; zdtint5; anti-angiogenic; intestinal polypeptide; wound healing;
 KW extracellular matrix interaction; tumour suppression; gamete maturation;
 KW immunologic recognition; gastrointestinal irradiation; chemotherapy;
 KW proteolysis; apoptosis; angiogenesis; infection; cell adhesion;
 KW cell fusion; cell signalling; tumour; Crohn's disease; melanoma; trauma;
 KW inflammatory bowel disease; food poisoning; degenerative disease;
 KW inflammation; fertility; gamete maturation; epithelial disorder.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 172
 FT /note= "potential N-linked glycosylation site"
 FT Modified-site 176
 FT /note= "potential N-linked glycosylation site"
 FT Modified-site 485
 FT /note= "potential N-linked glycosylation site"
 FT Modified-site 533
 FT /note= "potential N-linked glycosylation site"
 FT Modified-site 560
 FT /note= "potential N-linked glycosylation site"
 FT Modified-site 595
 FT /note= "potential N-linked glycosylation site"
 FT Modified-site 635
 FT /note= "potential N-linked glycosylation site"
 FT Misc-difference 474
 FT /note= "unspecified residue encoded by TTN"

XX WO200159112-A1.

XX 16-AUG-2001.

XX 09-FEB-2001; 2001WO-US04198.

XX 10-FEB-2000; 2000US-0501806.

PA

XX (ZYMO) ZYMOGENETICS INC.

XX Holloway JL, Sheppard PO;

XX WPI: 2001-522477/57.

XX N-PSDB; AAH74765.

XX New anti-angiogenic intestinal polypeptides, zdtint5 polypeptides, which
 PT are members of disintegrin proteases, for modulating extracellular
 PT matrix interaction, tumour suppression and wound healing -

XX Claim 2; Page 84-88; 92pp; English.

XX The present sequence represents a human zdtint5 polypeptide. The zdtint5
 CC polypeptide is an anti-angiogenic intestinal polypeptide. Zdtint5 is
 CC used for modulating extracellular matrix interactions. Zdtint5
 CC of polypeptides. Zdtint5 polynucleotides are useful as probes or primers
 CC to clone 5' non-coding regions of zdtint5 gene. Zdtint5 polypeptides are
 CC used for tumour suppression, gamete maturation, immunologic recognition,
 CC and growth and differentiation either working in isolation or in
 CC conjunction with other molecules in colon, small intestine, fetal lung,
 CC testis and B-cells. zdtint5 polypeptides are also useful for promoting
 CC wound healing, in the treatment of disorders associated with recovery
 CC after gastrointestinal irradiation, chemotherapy or antibody use.
 CC as anti-infectives, and extracellular matrix repair and remodeling. The
 CC polypeptides are also useful for modulating proteolysis, apoptosis,
 CC angiogenesis, infection, cell adhesion, cell fusion and signalling.
 CC The polypeptides are also useful for treating tumour formation, Crohn's
 CC disease, inflammatory bowel disease, food poisoning, melanoma, inflammation,
 CC fertility, gamete maturation, immunology, trauma and epithelial
 CC disorders.

XX Sequence 1120 AA;

Query Match 100.0%; Score 57; DB 22; Length 1120;
 Best Local Similarity 100.0%; Pred. No. 0.12;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILHLELLV 12
 DB 105 AAGGILHLELLV 116
 |||||

RESULT 9

AAG24449
 ID AAG24449 standard; Protein; 1353 AA.

XX AAG24449;

XX 04-OCT-2002 (first entry)

XX Human Von Willebrand factor-cleaving protease fragment #2.

XX Human; Von Willebrand factor-cleaving protease; vWF-cp; therapy; enzyme;
 KW transgenic animal; immunisation; thromboembolic disease; pre-eclampsia;
 KW thrombotic thrombocytopenic purpura; TTP; Henoch-Schönlein purpura;
 KW thrombosis; neonatal thrombocytopenia; haemolytic-uraemic syndrome;
 KW transgenic; anticoagulant.

XX Homo sapiens.

XX WO200242441-A2.

XX 30-MAY-2002.

XX 20-NOV-2001; 2001WO-EP13391.

XX 22-NOV-2000; 2000US-0721254.

XX 12-APR-2001; 2001US-0833328.

XX (BAXT) BAXTER AG.

XX Laemmle B, Gerritsen HE, Furlan M, Turecek P, Schwarz H;
 PI Scheiflinger F, Antoine G, Kerschbaumer R, Tagliavacca L;
 PI Zimmermann K, Voelkel D;
 XX WPI: 2002-479950/51.
 DR Novel isolated or substantially purified Von Willebrand factor-cleaving
 XX protease, useful for producing preparation for therapy of thrombosis
 PT and thromboembolic disease such as thrombotic thrombocytic purpura -
 XX
 PS Claim 1; Page 64-68; 93pp; English.
 XX The invention relates to an isolated or substantially pure Von Willebrand
 CC factor-cleaving protease (vWF-cp) polypeptide. vWF-cp is useful for
 CC purifying vWF which involves providing vWF-cp as a ligand, contacting a
 CC solution comprising vWF with the polypeptide ligand under conditions
 CC where vWF is bound to the ligand and recovering from the ligand purified
 CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
 CC which involves immunising an animal with vWF-cp and isolating the anti-
 CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
 CC producing a preparation of prophylaxis and therapy of thrombosis and
 CC thromboembolic disease such as thrombotic thrombocytic purpura (TTP),
 CC Hensch-Schönlein purpura, preclampsia, neonatal thrombocytopenia or
 CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
 CC plasmatric or recombinantly produced vWF. The invention is useful for
 CC construction expression systems and generating transgenic animals which
 CC express the polypeptide in vivo. The present sequence is human vWF-cp
 CC fragment.
 XX Sequence 1353 AA;
 SQ Query Match 100.0%; Score 57; DB 23; Length 1353;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAGGILHLELLV 12
 DB 1 AAGGILHLELLV 12
 RESULT 10
 ID ABB04153 standard; protein; 1416 AA.
 AC ABB04153;
 XX 26-MAR-2002 (first entry)
 DT Human ADAMTS-M polypeptide.
 DE Osteoarthritis; rheumatoid arthritis; inflammatory bowel disease;
 KW Crohn's disease; asthma; Alzheimer's disease; organ transplant rejection;
 KW cachexia; allergy; cancer; leukaemia; lymphoma; osteoporosis;
 KW atherosclerosis; congestive heart failure; myocardial infarction; stroke;
 KW neurodegenerative disease; autoimmune disorder; Huntington's;
 KW Parkinson's; migraine; pain; depression; multiple sclerosis; burn;
 KW infertility; diabetic shock; gene therapy; ADAMTS-M;
 KW A Disintegrin And Metalloprotease; thrombospondin domain.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH Peptide 1..97
 FT /label= Prodomain
 FT /note= "The mature form of the ADAMTS-M protein is
 FT processed by furin cleavage of the prodomain"
 FT 94..97
 FT Cleavage-site /label= Furin_cleavage_site
 FT 98..1416
 FT Protein /label= Mature ADAMTS-M protein
 FT /note= "The mature form of the ADAMTS-M protein is
 FT processed by furin cleavage of the prodomain"
 FT

FT Domain 98..311
 FT /label= Metalloprotease_domain
 FT Domain 247..272
 FT /label= Zinc-binding_motif
 FT Domain 324..394
 FT /label= Disintegrin_domain
 FT Domain 410..473
 FT /label= Thrombospondin_submotif
 FT Domain 419..424
 FT /label= Heparin-binding_domain
 FT Domain 1099..1156
 FT /label= Thrombospondin_submotif
 XX
 PN EP1152055-A1.
 XX 07-NOV-2001.
 PD
 XX 24-APR-2001; 2001EP-0303706.
 PF
 XX 27-APR-2000; 2000US-200040P.
 PR (PF12) PFIZER PROD INC.
 XX Buckbinder L, Mitchell PG, Wachtmann TS, Walsh RT;
 XX WPI: 2002-084275/12.
 DR N-PSDB; ABA02549.
 XX New polynucleotide, useful in gene therapy, particularly for treating
 PT or preventing e.g. arthritis, Crohn's disease, Alzheimer's disease and
 PT organ transplant toxicity and rejection, comprises ADAMTS
 PT polynucleotide and encoded polypeptide -
 PT
 XX Claim 4; Fig 2; 31pp; English.
 XX The present sequence represents a ADAMTS protein, designated ADAMTS-M,
 CC that exhibits the characteristics of the ADAM (A Disintegrin And
 CC Metalloprotease) family of metalloproteases, and contains a
 CC thrombospondin domain (TS). The protein is encoded by the cDNA given in
 CC ABA02549. The specification describes a newly isolated polynucleotide,
 CC comprising a nucleotide sequence encoding an ADAMTS-M polypeptide as
 CC given in the specification, or a metalloprotease, disintegrin domain,
 CC prodomain or its thrombospondin submotif. The polynucleotide, polypeptide
 CC and agent are useful for manufacturing a medicament for treating a
 CC subject in need of altering activity or expression of ADAMTS-M. The
 CC polynucleotide, ADAMTS-M polypeptide and agent are useful for
 CC manufacturing a medicament for treating arthritis (osteoarthritis and
 CC rheumatoid arthritis), inflammatory bowel disease, Crohn's disease,
 CC asthma, Alzheimer's disease, organ transplant toxicity and rejection,
 CC cachexia, allergy, cancer (e.g. solid tumour cancer including colon,
 CC breast, lung, prostate, brain or haematopoietic malignancies including
 CC leukaemia and lymphoma), osteoporosis, atherosclerosis, aortic aneurysm,
 CC congestive heart failure, myocardial infarction, stroke, head trauma,
 CC spinal cord injury, neurodegenerative disease, autoimmune disorders,
 CC Huntington's disease, Parkinson's disease, migraine, pain, depression,
 CC multiple sclerosis, abnormal wound healing, burns, infertility or
 CC diabetic shock. The polynucleotide and polypeptide are also useful for
 CC diagnosing the diseases above. The polynucleotide is particularly useful
 CC in gene therapy for treating the diseases cited above.
 XX
 SQ Sequence 1416 AA;
 Query Match 100.0%; Score 57; DB 23; Length 1416;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAGGILHLELLV 12
 DB 98 AAGGILHLELLV 109
 RESULT 11
 AAE24450

CC vWF. vWF-cp is useful for producing anti-vWF cp polypeptide antibodies
CC which involves immunising an animal with vWF-cp and isolating the anti-
CC vWF cp polypeptide antibodies from the animal. vWF-cp is useful for
CC producing a preparation of prophylaxis and therapy of thrombosis and
CC thromboembolic disease such as thrombotic thrombocytopenia or
CC Henoch-Schönlein purpura, preeclampsia, neonatal thrombocytopenia or
CC haemolytic-uraemic syndrome. vWF-cp can also be used for processing
CC plasmatic or recombinantly produced vWF. The invention is useful for
CC construction expression systems and generating transgenic animals which
CC express the polypeptide in vivo. The present sequence is human vWF-cp
CC N-terminal peptide.
XX
SQ Sequence 9 AA;
Query Match 78.9%; Score 45; DB 23; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGGILHLE 9
Db 1 AAGGILHLE 9
RESULT 13
ID ABG11310 standard; Protein; 669 AA.
XX
AC ABG11310;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #11301.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS75497.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 41669; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 669 AA;
Query Match 73.7%; Score 42; DB 22; Length 669;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy 1 AAGGILHLELLV 12
Db 10 AGGGLHLTL 21
RESULT 14
ID AAB42910 standard; Protein; 694 AA.
XX
AC AAB42910;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2674 polypeptide sequence SEQ ID NO:5348.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antihemetic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
DR N-PSDB; AAC77119.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11; Page 4526-4527; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397.

CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiseptatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antichyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergic, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SQ Sequence 694 AA;

Query Match 73.7%; Score 42; DB 21; Length 694;
 Best Local Similarity 66.7%; Pred. No. 37;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGGILHLELLV 12
 Db 10 AGGGLHLTLLL 21
 |||:|||||
 |||:|||||

RESULT 15

ABG11309
 ID ABG11309 standard; Protein; 694 AA.

XX AC ABG11309;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #11300.

XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS75496.

XX FT New isolated polynucleotide and encoded polypeptides, useful in
 XX FT diagnostics, forensics, gene mapping, identification of mutations
 XX FT responsible for genetic disorders or other traits and to assess
 XX FT biodiversity

XX PS Claim 20; SEQ ID NO 41668; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
 XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG0377 represent novel human
 CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 694 AA;

Query Match 73.7%; Score 42; DB 22; Length 694;
 Best Local Similarity 66.7%; Pred. No. 37;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGGILHLELLV 12

Db 10 AGGGLHLTLLL 21
 |||:|||||
 |||:|||||

Search completed: March 20, 2003, 12:33:25
 Job time : 9.24324 secs

CentCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:36:13 ; Search time 19 Seconds
(without alignments)
748.837 Million cell updates/sec

Title: SEQID_1_15FUSED
Perfect score: 773
Sequence: 1 AGGGLHLELLVAVGPDVQ.....SPTWSCLITEDTGFDLGVTI 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.0

Searched: 283224 seqs, 96134422 residues 283224

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	272	35.2	26926	1 I38344	titin, cardiac muscle
2	264	34.2	15281	2 S41309	cyclosporin synthet
3	244	31.6	10797	2 T30192	probable peptide s
4	238	30.8	10223	2 T30225	polyketide synthas
5	237	30.7	9376	2 T14593	syringomycin synth
6	233	30.1	7829	2 T15789	hypothetical prote
7	229	29.6	2165	2 T21371	hypothetical prote
8	229	29.6	7463	2 T36248	CDA peptide synth
9	225	29.1	7160	2 T27935	hypothetical prote
10	224	29.0	5255	2 T31677	bacitracin synthet
11	224	29.0	7576	2 T17428	PK506 polyketide s
12	222	28.7	5232	2 A45086	HC-toxin synthetas
13	221	28.6	6831	2 A88852	protein unc-22 (im
14	221	28.6	6839	2 S57242	twintin (similari
15	220	28.5	6420	2 T30283	polyketide synthas
16	220	28.5	8563	2 T30226	polyketide synthas
17	219	28.3	13055	2 T16580	hypothetical prote
18	217	28.1	6805	2 S20901	titin - rabbit (fr
19	217	28.1	8243	2 T31307	type I fatty acid
20	216	27.9	4845	2 T31067	BIR repeat contain
21	214	27.7	6260	2 T30228	polyketide synthas
22	212	27.4	5035	1 I46646	ryanodine receptor
23	211	27.3	4351	2 T00252	MEGF1 protein - ra
24	211	27.3	6658	2 T13931	projectin - fruit
25	210	27.2	5032	1 A35041	ryanodine receptor
26	210	27.2	5825	2 T12117	polyprotein - fava
27	210	27.2	6486	2 T31076	tyrocidine synthet
28	209	27.0	3649	1 S18268	delta-(L-alpha-ami
29	209	27.0	4836	2 T14346	herc2 protein - mo

30	209	27.0	5147	1 IJFFTM	cadherin-related t
31	207	26.8	4613	2 T17409	polyketide synthas
32	206	26.6	4302	2 A38971	polycystic kidney
33	206	26.6	4969	2 A37113	ryanodine receptor
34	206	26.6	5037	1 A34161	ryanodine-binding
35	205	26.5	4639	1 A54794	dynenin heavy chain
36	204	26.4	5005	2 F82884	hypothetical prote
37	204	26.4	5037	2 B35041	ryanodine receptor
38	203	26.3	4967	2 S72269	ryanodine receptor
39	203	26.3	5107	2 T29144	partial CDS - Caen
40	203	26.3	5175	2 T20992	hypothetical prote
41	203	26.3	5198	2 T43290	hemiscentin precurs
42	203	26.3	5369	2 T44807	mycosubtilin synth
43	203	26.3	6359	2 T31679	bacitracin synthet
44	203	26.3	6642	2 T29757	protein UNC-89 - C
45	202	26.1	3988	1 GNWVBV	genome polypeptide

ALIGNMENTS

RESULT 1

I38344
titin, cardiac muscle [validated] - human
N:Alternate names: connectin
N:Contains: serine/threonine-specific protein kinase (EC 2.7.1.-)
C:Species: Homo sapiens (man)
C:Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 15-Sep-2000
C:Accession: I38344; I38345; S20898; S20897; S20899; S63665; S37393
R:Labelt, S.; Kolmerer, B.
S:Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330; PMID:7569978
A:Accession: I38344
A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/
A:Molecule type: mRNA
A:Residues: 1-26926 <LAB1>
A:Cross-references: EMBL:X90568; NID:g1017424; PID:g1017425
R:Musco, G.; Tlatziolos, C.; Schuck, P.; Pastore, A.
B:Biochemistry 34, 553-561, 1995
A:Title: Dissecting titin into its structural motifs: identification of an alpha-heli
A:Reference number: I38345; MUID:95119041; PMID:7819249
A:Accession: I38345
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1977-2014 <MUS>
A:Cross-references: EMBL:X83270; NID:9602579; PIDN:CAA59243.1; PID:9602580
A:Note: Conformation and properties are reported for a synthetic peptide correspondin
R:Labelt, S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A:Title: Towards a molecular understanding of titin.
A:Reference number: S20897; MUID:92258380; PMID:1582406
A:Accession: S20898
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 13597-14200, '1', 14202-14696 <LAB2>
A:Cross-references: EMBL:X64698; NID:g37192; PIDN:CAA45939.1; PID:g37193
A:Accession: S20897
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB3>
A:Cross-references: EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID:g37191
A:Accession: S20899
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'O', 22455-22480, 'TR', 2248
A:Cross-references: EMBL:X64697; NID:g37190; PIDN:CAA45938.1; PID:g37195
R:Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labelt, S.
J. Mol. Biol. 256, 556-563, 1996
A:Title: Genomic organization of M line titin and its tissue-specific expression in t
A:Reference number: S63665; MUID:96177761; PMID:8604138
A:Accession: S63665
A:Status: nucleic acid sequence not shown

A:Molecule type: DNA
A:Residues: 26729-26825 <KOL>
A:Cross-references: EMBL:X92412; NID:G1236761
R:Gautel, M.; Leonard, K.; Labelit, S.
EMBO J. 12, 3827-3834, 1993
A:Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentia
A:Reference number: S37393; MUID:94008990; PMID:8404852
A:Accession: S37393
A:Molecule type: mRNA
A:Residues: 26831-26926 <GAU>
R:Improt, S.; Politou, A.S.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, February 1996
A:Reference number: A66736; PDB:1TIT
A:Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
R:Fuhl, M.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A66201; PDB:1NCT
A:Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q32
C:Function:
A:Description: structural protein forming filaments in striated muscle
C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pro
C:Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco
structural protein
F:24752-25008/Domain: protein kinase homology <KIN>
F:84,177,905,2378,2459,2481,2563,2669,2763,2896,3088,3179,3384,3432,3628,3772,4069,
98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,13540,1
tatus predicted
F:16780,16976,17579,17602,17667,17681,17845,17899,18121,18188,18209,18336,18670,18680,18
,21900,21935,22295,22495,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,248
F:26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 35.2%; Score 272; DB 1; Length 26926;
Best Local Similarity 0.4%; Pred. No. 49;
Matches 101; Conservative 16; Mismatches 23; Indels 24887; Gaps 28;

Qy	9	ELL-----	11
Db	190	ELLVQGEVEPAKTKTIVSTAQISERQTRIEKIEAHFADARSATVEMVIDGAAGQOL	249
Qy	12	-----	11
Db	250	PHKTPPRIPPKRSRPTPPSIAAKAQLARQOSPIRHSPPVRHVRAPTSPVRSP	309
Qy	12	-----VAVGPDV-----	18
Db	310	AARISTPISRSVSPLLMRKTKQASTVATGPEVPPPKQEGYVASSSEAEEMRETTLTSTQ	369
Qy	19	-----	18
Db	370	INTEREGRVGVQVITISGAAGAAVSASASAAEAVATGAKEVKQADKSAAVATV	429
Qy	19	-----	18
Db	430	VAAVDMARVREPVISAVEQTAQRTTTTAVHIQPAQEQVRKEAEKTAVTAVVVVAADKAKEQ	489
Qy	19	-----	18
Db	490	ELKSRTKEIITTKQEQMHVTHEQIRKETKTFVPKVVISAAKAKEQETRISEITKKQKQ	549
Qy	19	-----	18
Db	550	VTOEAIMKETRTKTVPKVIVATPKVKQDLVSRGREGITTKREQVQITQEKMRKEAKTA	609
Qy	19	-----	18
Db	610	LSTIAVATAKAKEQETILTRTETMATRQEQIQVTHGKVDVGKAEAVATVVAADQARV	669
Qy	19	-----	18

Db	670	EPREPGHLEESYAQOTTLEYGYKERISAAKVAEPPOPPASEPHVVPKAVKPRVIOAPSET	729
Qy	19	-----	18
Db	730	HIKTTDKQGMHISSQIKKTTDLTTERLVHVKRPTASPHFTVSKISVPKTEHGYEASIA	789
Qy	19	-----	18
Db	790	GSAIATLQKELSATSQAQITKSVKAPTVPKSETRVRAEPTPLPQFPFADTPTDYKSEAG	849
Qy	19	-----	18
Db	850	VEYKEVGSITGTTVREERFEVLHGREAKVTETARVPAPVEIPVTPPTLVSLGNVTVI	909
Qy	19	-----FO-----AHQSDTERV-30	30
Db	910	EGESVTLECHISGYPSPPTVWYREDYQIESSIDFOITTFOSGIARLWIREAFEDSGRFTC	969
Qy	31	-----	30
Db	970	SAVNEAGTVSTSCYLAVQVSEBFEKETTAVTEKFTTEKRFVESRDVMTDTSLTTEEQAG	1029
Qy	31	-----	30
Db	1030	PGEPAAPYFITKPVOKLVEGGSVFGCOVGNPKPHVWKKSGVPLTTGTRYKYSYNKO	1089
Qy	31	-----	30
Db	1090	TGECKLVISMTFADDAGEYTIIVRNKHGETSASALLEEADYELLMKSOQEMLYQTVTA	1149
Qy	31	-----	30
Db	1150	FVQPEVGETAPGVFVSEVEKEYEQALIRKQAKDVTVVRTYVEDQEPHISSEERLI	1209
Qy	31	-----	30
Db	1210	KEISYRIITKLEELLEDGEKXAVDISESEAVESGDLRIKNYRILEGMVTFHCKMS	1269
Qy	31	-----	30
Db	1270	GYPKPIANYKDGKRIKHGERYOMDFLOGRASLRIPVLVPEDEGIYAFASNIKGNNAIC	1329
Qy	31	-----	30
Db	1330	SGKLYVEPAAPLGAPTYIPTLEVPVSRISLSPRSVSRSPIRMSPARMSPARMSPA	1389
Qy	31	-----	30
Db	1390	RMSGRRLEETDESQERLYKPVFLKPVSKFCLEGANCDFDLKVGRPMPTFWFHDGQ	1449
Qy	31	-----	30
Db	1450	QIVNDYTHKVIKEDGTOSLIIVPATPSDSGEWTVVAQNRAGRSSISVILTVEAVEHQV	1509
Qy	31	-----LTNLI-----	36
Db	1510	PMFVEKLVNVIKEGSRLEMKVRATGNPNPDIVMLKNSDIIVPHKYPKIRIEGTKGAAL	1569
Qy	37	-----	36
Db	1570	KIDSTVSQDSAWYTATAINKAGRDTTTRCKNVVEVEFAEPEPERKLIIPRGTYRAKEIAAP	1629
Qy	37	-----	36
Db	1630	ELEPLHURYGOEQWEGDLYDKEKQKQKFFKKLTSRLKRFKPAHFECRLTPISDPTMV	1689
Qy	37	-----	36
Db	1690	VEWLHDGKPLEAANRLRMINEFGYCSLDYGVAYSRSDSGIITCRATNKYGTDHTSATLIVK	1749
Qy	37	-----GA-----	38
Db	1750	DEKSLVBESQLPEGRKGLORIEELERMAHEGALTGVTTDOQEKOKPDIVLYPEPVRVLEG	1809

QY	39	-----	38	QY	59	-----	58
Db	1810	ETARFCRVGYPOPKVWYLNQILIRSKRFRVRYDGIHYLDIVDCSKSYDTGEVKVTAE	1869	Db	2890	HQLIMNTSTEDSAEYTFVCGNDQVSATLTVPIMITSMMLKDINAEKDTITFEVTNYE	2949
QY	39	-----	38	QY	59	-----	58
Db	1870	NPEGVIEHKVLEIOQREDFRSLRRAPEPRPEFHVHBPGLQFEVQKVDPRVDTTETKE	1929	Db	2950	GISYKWLKNGVEIKSTDKCQMRKTLTHSLNIRNVHFGDAADYTFVAGKATSTATLYVEA	3009
QY	39	-----	42	QY	59	-----	58
Db	1930	VVKLKRAERITHEKVPESEELRSKFRTEEGYYEAITAVELKSRKXDESVEELRTK	1989	Db	3010	RHIEFRKHIDIKVLEKKRAMFECEVSEPDITVQMKDDQEOLOITDRIKIOKEYVHRL	3069
QY	43	-----	42	QY	59	-----	58
Db	1990	DELLHWTKELTSEKKALAEBSKTIPTFKPKIELSPSMEAPKIFERIQSOTVQGSDA	2049	Db	3070	IPSTRMSDAGKYTVVAGGNVSTAKLFVEGRDVRIRSIKKEVQVIEKQRAVVEFVNEDDV	3129
QY	43	-----	45	QY	59	-----	58
Db	2050	HERVVRVGPDPCECEWYKNGVKIERSDRIYWYWPEDNVCELVRDVTAEADSASIMVKAIN	2109	Db	3130	DAHMYKDGIEINFQVQERHKYVVERIHRMPFISERQSDAGEYTFVAGNRSSVTLVNA	3189
QY	46	-----	45	QY	59	-----	58
Db	2110	IAGETSSHAFLVQAKLITFTQELQDVVAKEDTMATFECETSEPFVVKWYKDGMEVH	2169	Db	3190	PEPQVQLQELQVTVQSGKPARFCAMISGRPPKISWYKERQLLSTGFKCFLHDOBYT	3249
QY	46	-----	45	QY	59	-----	58
Db	2170	EGDKYRMHSDRKVHFLSILITDTSDAEDYSCVLVEDENVKTTAKLIVEGAVVEFVKELOD	2229	Db	3250	LLLIEAFPEDAAYTCEAKNDYGVATTSSASLSEVPEVSPDOEMPVYPPIITPLQDTV	3309
QY	46	-----	45	QY	59	-----	58
Db	2230	IEVPESYSGELECIVSPENIEGMYHNDVELKSNKYTITSRGRQNLTVKDVTKEDQGE	2289	Db	3310	TSEGOPARFQCRVSGTDLKVSWYKDKKIKPSRFRMTQPEDTYQLEIAEAYPEDEGTYT	3369
QY	46	-----	45	QY	59	-----	58
Db	2290	YSFVIDGKTKTKLKMPPRIATLOGLSDQKCEGDIVOLEVKVLSVESVEGVMKDGQEV	2349	Db	3370	FVANNAVQSVSTANLSLEAPESILHERIEQIEIEMEMKEPSSFLSABEEGLHSAEOLS	3429
QY	46	-----	45	QY	59	-----	58
Db	2350	QPSDRVHIVIDKOSHMLLIEDMTKEDAGNYSFTIPALGLSTGRVSVYSDVITPLKDVN	2409	Db	3430	KINETLELLSESPVPTKFDSEKGTGPIFIKEVSNADISMGDVATLSVTVIGIPKPKIQ	3489
QY	46	-----	45	QY	59	-----	58
Db	2410	VIEGTKAVLECKVSPDVTSVKWLNDQEIKPDDRQVAIVKGTQORLVINRTHASDEGPY	2469	Db	3490	WFFNGVLLTPSADYKVFDDGDDHSLIILFTKLEDEGEYTCMASNDYKGTICSAYLKINSK	3549
QY	46	-----	45	QY	59	-----	58
Db	2470	KLIVGRVETNCNLSEKIKIIRGLRDLTCTETQNVVFEVLSHSGIDVLWNFKDKKEIKPS	2529	Db	3550	GECHKDTETESAVAKSLEKLGCPHPFLKELKPIRCAQGLPAIPEYTVVGEPAPTVTFW	3609
QY	46	-----	45	QY	59	-----	58
Db	2530	SKYKIEAHGKIYKLTVLNMMKDDGKYTFYAGENMTSGKLTVAGGALSPLTDQTVAESQ	2589	Db	3610	KENKOLCTSVVYTIHNPNGSGTFIWDPOREDSGLYICKAENMLGESTCAELLVLLLED	3669
QY	46	-----	45	QY	59	-----	58
Db	2590	EAVFECEVANPSKGWLRDGHPLTNIRSESDGHRRLIIAATKLLDDIGEYTYKVAT	2649	Db	3670	TDMTDTPCKAKSTPEAPEDFPQTPPLKGPVAEALDSEQEIATEVKDTILKAALITEENOL	3729
QY	46	-----	45	QY	59	-----	58
Db	2650	SKTSAKLKVEAVKIKKTLKNTVTETODAVTVELTHPNVKGQWIKNGVVLSENEKYAI	2709	Db	3730	SYEHIAKANELSSQLPLGAQELQSILODKLTPSTREFLCINGSIHPOPLKEPSPNLOL	3789
QY	46	-----	58	QY	59	-----	65
Db	2710	SVKGTIYSLRIKCAIVDESIVYGRGLGASARLHVETVKIHKPKDVTALENATVAPE	2769	Db	3790	QIVSQKTFSEKILMPPEPETOAVLSOTEKIFPSAMSIEQINSLTVEPLKTLAEPEGN	3849
QY	59	-----	58	QY	66	-----	65
Db	2770	VSVSHDTPVKWFKHSVEIKPSDKHRLVSRKVHKLMLQNI SPSDAGEYAVVGQLECKA	2829	Db	3850	YQSSIEPPMHSVLTSAVEEVLSEKKTVSDTNREQRTVLOKQEAQSALILSOSLAEGHV	3909
QY	59	-----	58	QY	66	-----	65
Db	2830	KLFVETLHITKMKNIIEVPTKTATSECEVSHFNVPNSMNLKNGVEIEMSEKFIIVQOKL	2889	Db	3910	ESLQSPDVMISQVNEPLVPSHSECHTGGKILIESANPLENAGQDSAVRIEKGSLRFPPL	3969
				QY	66	-----	65

Thu Mar 20 13:50:29 2003

Db 3970 ALEKQVLLKEHSDNVMPDQIIESKREPVAIKKVEQVGRDLLSKESLLSGIPPEQR 4029
QY 66 ----- 65
Db 4030 LNLKIQICRALQAASEQPLFSEWLNIEKVEVEAVNITQEPHIMCMYLVTSKSVT 4089
QY 66 ----- 65
Db 4090 BEVTIIIEDVDPQMANLKMELRDALCAIYYEEDILTAEGRIOQGAKTSLQEEMDSFG 4149
QY 66 ----- 65
Db 4150 SQKVEPITEVESKYLISTEESYFNVQSRVKYLDATPVTKVASAVSDEKODESLKP 4209
QY 66 ----- 65
Db 4210 SEEKESSESSEGTAEVTVKIOEAGGLIKEDGPMIHTPLVDTVSEEGDIVHLTTSITNA 4269
QY 66 ----- 65
Db 4270 KEVNWYFENKLVPSDEKFKLQDQNTVTLVIDKNTEDHOGVEVCEALNDSGKTATSAL 4329
QY 66 ----- 69
Db 4330 TVWKRAAPVIRKRIEPLVALGHLAKFTCEIQSAPNVRFOFKAGREIYESDKCSIRSSK 4389
QY 70 ----- 69
Db 4390 YISLEILRTQVDCGEYTCASNEYSVSCTATLTVTVPGEKKVRKLLPERKPEKEE 4449
QY 70 ----- 69
Db 4450 VVLKSLRKRPEEBEKVEPKKLEKVPAPPEPPPKPVEEVPVTVTKRERKIPEPTK 4509
QY 70 ----- 69
Db 4510 VPEIKPAIPLPAPEPKPEAEVTKIKPPVPEPPTPIAAPVTVPVVKKAIAKAPKEEA 4569
QY 70 ----- 69
Db 4570 AKPGPVKVPKTPSPIEAERKLRPGSGGKPPDEAPFTYQLKAVPLKPVKBIKDIIL 4629
QY 70 ----- 69
Db 4630 TESEFVGSSAIFECVLSPSTAITTMMKDGSNIRESPKHFADGDKRKLHIIDVQLSDAG 4689
QY 70 ----- 69
Db 4690 EYTCVLRGNKEKSTAKLVVEELPVRFVKTLBEEVTVKGPPLYLSCELNKRERDVWRK 4749
QY 70 ----- 69
Db 4750 DGIIVVEKGRIVPGVIGLMRALTINDADDTAGTYTTVTENANNLECCSCVKVVEIRD 4809
QY 70 ----- 69
Db 4810 WLVKPIRQHVKPGKTAIFACDIAKDTNPKWFKGYDEIPAEPNDKTEILRDGNHLYLKI 4869
QY 70 ----- 69
Db 4870 KNAMPEDIAEYAVEIEGKRYPAKTLGERVELLXPIEDVTIYEKESAFDAEISEADIP 4929
QY 70 ----- 69
Db 4930 GQWKLKGELLRSPPTCEIKAEGGRFLTLHKVLDQAGEVLYQALNAITTAILTVKEIEL 4989
QY 70 ----- 69
Db 4990 DPAVPLKDVTPERRQARFECVLTREANVWSKGPDIKSDKFDIIADGKHHILVINDS 5049
QY 70 ----- 69

Db 5050 QPDDGVTYAEVEGKKT SARLFTVTGIRLKFMSPLEDQTVKEGETATFVCELSHERKHVVW 5109
QY 70 ----- 69
Db 5110 FKNDAKLHTRTSLVLISSSEKTHKLEMKEVTLDDISQIKAQVKELSSSTAOLKVLEADPYFT 5169
QY 70 ----- 69
Db 5170 VKLHDKTAVEKDEITLKEVSKDVPVKWFKDGEIIVPSPKYSIKADGLRRLIKKADLJK 5229
QY 70 ----- 69
Db 5230 DKGEYVDCDGTDKTKANVTVEARLIEVEKPLYGVFVGETAHFETELSEPDVHGOMKLU 5289
QY 70 ----- 69
Db 5290 GQPLTASPDCEIIEBGKXHLILHNCOLGTMTEVSGFOANAKSAANLKVKEPLFITPL 5349
QY 70 ----- 69
Db 5350 SDVKVPEKDEAKPECEVSESRPKTFRWLKGTQEIITGDRFELIKDGTKHSWIKSAAFEDE 5409
QY 70 ----- 69
Db 5410 AKYMPAEADKHTSGKLIIEGIRLKLTLPLKDVTAKEKESAVFTVELSHONIRVKWFKNDQ 5469
QY 70 ----- 69
Db 5470 RLHTRSVMOEGKTHSITFKDLSIDDTSQIRVENAGMSSSEAKLTVLSEDPYFTCKLQD 5529
QY 70 ----- 69
Db 5530 YTCVEKDEVILQCEISKADAPVKWFKDGEIKPSKNAVIKTDGKKRMLIKKALKSDIGQ 5589
QY 70 ----- 69
Db 5590 YTCDCGTDKTSGKLDIEDREIKLVRPLHVSVMETETARFETIEISDDIHANWKLGEAL 5649
QY 70 ----- 69
Db 5650 LQTPDCEIKEEGKIHSLVHLNCRLDQDGTGVDFOANVKSASHLVRVKPRVIGLLRPLKDV 5709
QY 70 ----- 69
Db 5710 VTAGEATDFCELSYEDI PVWYLKGGKLEPSDKVVPSEGKVHTLTLDVKLEDAVEQ 5769
QY 70 ----- 69
Db 5770 LTAKDFKTHANLFVKEPPVEFTKPLEDQTVVEGATAVLECEVSRENAKVWFKNGTEILK 5829
QY 70 ----- 69
Db 5830 SKKYEIVADGRVRKLVHDCPTEDIKTYTCDKDFKTSNVLNVPPHVEFLRPLTDLQVR 5889
QY 70 ----- 69
Db 5890 EKEMARPECELSRENKVKWFKDGEIKGKYDIISKGAVRILVINKCLLDDDEAEYSCE 5949
QY 70 ----- 69
Db 5950 VRTARTSGMLTVLEEEAVFTKNLANIEVSETDTIKLVCEVSKPGAEVWYKGDDEIIEGT 6009
QY 70 ----- 69
Db 6010 RYEILTEGRKRLVIONAHLEDAGNYNCRLPSSRTDCKVKVHLEAAEFISKPNLEILEG 6069
QY 70 ----- 69
Db 6070 EKAEFVCSISKESPPVQWKRDDKLTLESGDKYDVIADGKRVLVVYKATLQDMGTVVWVG 6129
QY 70 ----- 69
Db 6130 AARAAHLTVIEKLRIVVPLKOTRVKEQQEVVFNCEVNTGAKAKWFRNEEAI FDSKKYI 6189

```
QY 70 ----- 69
Db 6190 ILQKOLVYTLIRDAHLDDQANYNSLTNRGENVKSAAANLIVEBEDLRIVEPLKDIETM 6249
QY 70 ----- 69
Db 6250 EKKSVTFWCKVNLNVTLKWTKNGBEVPFDNRVSYRVDKYKMLTIKDCGFPDEGEYIVT 6309
QY 70 ----- 69
Db 6310 AGODKSAELLIIIEAPTEFVEHLEDQTVTEFDDAVFCQLSREKANVWYRNGREIKEGK 6369
QY 70 ----- 69
Db 6370 KYKFEKDSIHLRII IKDRLDDECEYACGVDRKSRARLFVEEIPVEIIRPPQDILEAPG 6429
QY 70 ----- 69
Db 6430 ADVVFLAELNKKVEVQWLRNNMVVQGDQKQWSEGIHRLQICDIKPRDQGEYRFIAK 6489
QY 70 ----- 69
Db 6490 DKEARAKLELAAPKIKTADQDLVDVGKPLTWVPYDAYPKAEAEWFKENEPLSTKTIID 6549
QY 70 ----- 69
Db 6550 TTAQTSFRILEAKGDKGRYKIVLQKHGKAEGFINLKVIVDPGVPVRLVETETFDGEV 6609
QY 70 ----- 69
Db 6610 SLAWEPLTDGSKIIGYVVVERRDIKRKTWVLATDRAESCEFTVTGLQGGVEYLFVSA 6669
QY 70 ----- 69
Db 6670 RNRVGTGEPVETDNPVEARSKYDVPGLPLNVTITDVRFGVSLTWPEPPDYDGAETNYV 6729
QY 70 ----- 69
Db 6730 IELRDKTSIRWDAMTVRAEDLSATVTDVVEGQESYRVRQAQNRIGVGKPSAATPFVKVA 6789
QY 70 ----- 76
Db 6790 DPTERPSPVNLTSDDQTSQSVQLKWEPLKDGSPILGVIIEBCBEGKDMIRCNMKLV 6849
QY 77 ----- 76
Db 6850 PELTYKVTGLEKGNKYLYRVAENKAGVSDPSEILGPLTADDAFVEPTMDLSAFKDGLEV 6909
QY 77 ----- 76
Db 6910 IVNPITILVPSTGYPRPTATWCFGDKVLETGDRVKMKTLSAYAELVISPSERSDKGIYT 6969
QY 77 ----- 76
Db 6970 LKLENRVKTIISGEIDNVNVIARPSAPKELKFGDITKDSVHLTWEPDDDDGSGPLTYVVEK 7029
QY 77 ----- 81
Db 7030 REVSRTWTKVMDFVTDLEFTVPDLVOGKEYLFVKYCARNKGPGEPAYVDEPNMSTPAT 7089
QY 82 ----- 81
Db 7090 VPDPPENVKWRDRTANSIFLTWDPKNDGGSRIKGYIVERCPRGSKDWACGEPVAETKM 7149
QY 82 ----- 81
Db 7150 EVTGLEEKWYAYRVKTLNROGASKPSRPTBEIQAVDTQEAPEIFLDVKLLAGLTVKAGT 7209
QY 82 ----- 81
Db 7210 KIELPATVTKPEPKITWTAKMILKODKRITIEVNPVKSTVTIVDSKRSOTGTGYIIEAV 7269
```

```
QY 82 ----- 81
Db 7270 NVCGRATAVBEVNVLDKPGPAAFDITDVTNESCLLTWNPPRDDGSKITNYVVERRATD 7329
QY 82 ----- 81
Db 7330 SEVMHKLJSTVKDTNFKATKLIIPNKEYIFRVAENMYGAGEPVQASPIYAKYQDFPPGPP 7389
QY 82 ----- 81
Db 7390 TRLEPSDITKDAVTLTWCEPDDDGSPITGYWVERLDPDTKWVRCKMPVKDITTYRVKG 7449
QY 82 ----- 81
Db 7450 LTNKKYRFRVLAENLAGPGKPSKSTEPILIKDIPDPMPFGKPTVKDVGKTSVRLNWK 7509
QY 82 ----- 81
Db 7510 PEHGGAKIESYIEMLKTGTDEWVRVAEGVPTTQHLLPLGLMEGOEYSFRVRANVKAGES 7569
QY 82 ----- 81
Db 7570 EPSESDPVLCKEKLPPSPPPRWLEVINITKNTADLKWTVPEKDGSPITNYVIVEKRDVR 7629
QY 82 ----- 81
Db 7630 RKGQVTDVTKTKTCTVPLTEGSLYVFRVAENAIQGSYTEIEDSVLAKDFTTTPGP 7689
QY 82 ----- 81
Db 7690 PYALAVDVTKRHVDLKWEPKNDGGRPIQRYVIEKKERLGRVWVAKGTAGPCDNFRVT 7749
QY 82 ----- 81
Db 7750 DVIETGEVQFQVRAENEAGVHPSEPTIELSIEDPTSPSPPLDLHVTDAGRKHIAIAWK 7809
QY 82 ----- 81
Db 7810 PPEKNGSGPIIGYHVEMCPVGTEKMMRVNSRP IKDKFKEGVVVDKEYYLVRVAENAI 7869
QY 82 ----- 81
Db 7870 GYSEPEISENVVAKDPCKPTIDLETHDII VIEGEKLSIPVPPRAVPVPTVSWHKDGKE 7929
QY 82 ----- 81
Db 7930 VKASDLTMKNDHISAHLEVPKSVRADAGIYITITLENKLSATASINVKVIGLPGCKDI 7989
QY 82 ----- 81
Db 7990 KASDITKSSCKLTWBPFPDGGTPIHYVLERREAGRRTYIPVMSGENKLSWTVKDLIPN 8049
QY 82 ----- 81
Db 8050 GEYFFRVKAVNKVGGEYIELKNPVIAQDPKQPPDPVVDVEVHNPTAEAMTITWKPLD 8109
QY 82 ----- 81
Db 8110 GGSKINGYIIIEKIAKEERWKRNEHLVPLITYAKGLEGEKGYOFRVRAENAAAGISEPS 8169
QY 82 ----- 81
Db 8170 RATPPTKAVDPIDAPKVILRTSLEVRGDEIAIDASISGSPYTTITWIKENVIVPEEIK 8229
QY 82 ----- 81
Db 8230 KRAAPLVRRRKGEVQEEFPVLPTQRLSDNSKKESQLRVDRSLRPHGLYMIKIVEND 8289
QY 82 ----- 81
Db 8290 HGIAPACTVSLDTPGPPINFVEDIRKTSVLCKWEPLDDGGSEIINVTLEKKDKTKP 8349
QY 82 ----- 81
```

Db	8350	DSEWVVTSTLRHCKYSVTKLIEGKEYLFRVRAENRFGPPGCVSKPLVAKDPFGPDAP	8409
Qy	82	-----	81
Db	8410	DKPIVEDVTSNMLVKWNEPKDNGSPILGYLWLEKREVNSTHWSRVNKSLLNALKANVDGL	8469
Qy	82	-----	81
Db	8470	LEGLTYVFRVCAENAGPGKFPSPDPKTAHDPISPGPPPIPRVTDTSSTTIELEWEPPA	8529
Qy	82	-----	81
Db	8530	FNGGEIVGVFDKQLVGTNKNRSCTEKMKVROYTVKEIREGADYKLRSVANAAGEP	8589
Qy	82	-----	81
Db	8590	PGETQPVTVAEPPQPPAVELDVSVKGGIOMAGKTLRIPAVVTGRPVPTKVWTKKEGELD	8649
Qy	82	-----	81
Db	8650	KDRVIDNVGTSKSELIIDALRKDHGRVITATNSCGSKFAAARVEFVPGPVLDPV	8709
Qy	82	-----	81
Db	8710	VTNRKMLNWDPEDDGGSEITGFIIERKDAKMHTWRQPIETERSKCDITGLLEGOEYK	8769
Qy	82	-----	81
Db	8770	PRVIANKFGCGPPVEIGPILAVDPLGPPTSPERLTYTERORSTITLDWKEPRNSGSGPI	8829
Qy	82	-----	81
Db	8830	QGYIIEKRRHKDPFERVNRKLCPTTSPLENLDEHOMYEFYRVKAVNEIGESEPSLPV	8889
Qy	82	-----	81
Db	8890	VIQDEVPPTIKLRSVRGDTIKVAGEPVHI PADVTGLPMPKIEMSKNETVIEKPTDAL	8949
Qy	82	-----	81
Db	8950	QITKEEVSRSKTELSTIPKAVREDKGTYYVTASNRLGSVFRNVHVEYVDRSPRNLAV	9009
Qy	82	-----	81
Db	9010	TDIRAESCYLETWDAPLDNGGSEITHYVIDKRDSRKAEBEVTNTAVEKRYGIWKLIPN	9069
Qy	82	-----	81
Db	9070	GQYEFVRVANKYGISDECKSDKVVIQDPYRLPGPPGPKVLARTKGSMLVSWTPPLDNG	9129
Qy	82	-----	81
Db	9130	GSPITGYWLEKREEGSPYWSRVSRAPITKVGLKGVEFNVPRLLEGVKYQFRAMAINAGI	9189
Qy	82	-----CW-----	83
Db	9190	GPPSEPSDEVAGDPIFPFGPPSCPEVKDKTKSSISLGMKPPAKDGGSPIKGYIVEMOEE	9249
Qy	84	-----	83
Db	9250	GTTDWKRVNPDKLITTCCEVVPNLKELRKYFRVKA VNEAGESEPSDTTGEIPATDIOE	9309
Qy	84	-----	83
Db	9310	EPEVFIDIGAQCLVKAGSQIRIPAVIKGRPTPKSWEFFDGKAMKMGVHDIPEDAQ	9369
Qy	84	-----	83
Db	9370	LETAENSSVIIIECKRSHTGYSITAKNAQCKTANCRVKVMYVPGPPKDLKVSIDITRG	9429
Qy	84	-----	83
Db	9430	SCRLSWKMPDDGGDRIKGYVIEKRTIDGKAWTKVNPDCGSTTFVVPDLLSEQOYFRRVR	9489
Qy	84	-----	83
Db	9490	AENRFIGGPPVETIORTTARDPIYPPDPPIKLIGLITKNTVHLSWKPPKNDGGSPVTHY	9549
Qy	84	-----	83
Db	9550	IVECLAWDPTGKKEAWROCNRDVEELOFTVEDLVEGGEYEFYRVKANVNAAGVSKPSATV	9609
Qy	84	-----	83
Db	9610	GPCDCORPMPSPIDLKEPMEVEEGTNVNI VAKIKGVPPTLTWFKAPPKPDNKEPVLY	9669
Qy	84	-----	83
Db	9670	DTHVNKLAVDDTCTLVIQSRSDTGLYTTITAVNNLGTASKEMRLNLVLRGPPVGPPIKF	9729
Qy	84	-----	83
Db	9730	ESVSADQMTLSWPPKDDGGSKITNYVIEKREANRKTWHVVSSEPKECTYTIPTKLEHGE	9789
Qy	84	-----SOTIN-----	88
Db	9790	YVFRIMAQNKYGIGEPDSEPETARNLFSVPGAPDKPTVSSVTRNSMTVNWEPEYDGG	9849
Qy	89	-----	88
Db	9850	PVTGYWLEMKDTTSKRWKRVNRDPIKAMTLGVSVKVTGLIEGSDYOFRVVAINAAGVPA	9909
Qy	89	-----	88
Db	9910	SLPSDPATARDPIAPGPPPKVTDMTKSADLEWSPPLKDGSKVTGYIYVEKESKEE	9969
Qy	89	-----	88
Db	9970	WEXGKDEVRGTVLAVTGLKEGAFYKFRVSAVNIAGIGEPGEVTDVIEMKORLVSPLQL	10029
Qy	89	-----	88
Db	10030	DASVRDRIVHAGGVIIRIAYVSGKPPPTVTWMNERTLPOEATIIETAISSSMVINKCO	10089
Qy	89	-----PEDD-----	92
Db	10090	RSHQGVYLLAKNEAGERKKTIIVDVLDVPGPVGTPPLAHNLTNESCKLTFWSPEDDGG	10149
Qy	93	-----	92
Db	10150	PITNYVIEKRESRRAWTPVYTVTRONATVOGLIOGKAYFFRIAAENSIGMGFFVETSE	10209
Qy	93	-----	92
Db	10210	ALVIREPITVPERPEDLEVKEVTQNTVTLTNBPKYDGGSEIINYVLESRLIGTEKPHKV	10269
Qy	93	-----	92
Db	10270	TNDNLSRKYTVKLGEGDTEYRVSAVINVGOKPSFCTKPTCKDELAPTLHLDFRD	10329
Qy	93	-----	92
Db	10330	KLATIRGEAFALTCRYSGKPKPKVSWFKOBADVLEDDRTHIKTTPATLALEKIKAKRSDS	10389
Qy	93	-----	92
Db	10390	GKYCVVENSTGSRKGCQVNVVDHPGPPVGPVSFDEVTKDYMVISWKPLDDGSGKITN	10449
Qy	93	-----	92
Db	10450	YIIEKVEGKDVMPVTSASAKTTCKVSKLEGGDYIFRIHAENLYGISDPLVSDSMKAK	10509
Qy	93	-----	92
Db	10510	DRFRVPADQPIVTEVTKDSALVTWNKPHDGGKPIITNYILEKRETMSKRWARTKDPH	10569

QY	93	-----92	QY	102	-----101
Db	10570	PYTKFRVDPDLEGCOYEFVRVSAENEIGIGDPPSPKPFVAKDPIAKPSPVNNPEAIDTTC	Db	11650	WERVHKGSIKETHYMDRCVENQIYEFRTQNEGGESDMWKTBEVVVVKEDLQKPVLDLK
QY	93	-----92	QY	102	-----101
Db	10630	NSVDLTWQPRHDGSKILGYVEYQKVGDEBRRANHTPESCETKYKVTGLRGDTYK	Db	11710	LSGVLTKAGDTIRLEAGVRGKPFPEVAVWTKDADTLTRSPRVKIDTRADSSKFSLTKA
QY	93	-----97	QY	102	-----101
Db	10690	FRVLAVNAAGESDPAHVPEFLVKORLEPPELILDANMAREQHIKVGDTLRLSAIIKGV	Db	11770	KESDGGKYVVTATNTAGSFVAYATVNVLDKPGVVRNLKIVDVSSDRCTVCWDPEDDGGC
QY	98	-----97	QY	102	-----101
Db	10750	FPKVTWKEDRDAPTKARIDVTPVSKLEIRNAAHEDGGIYSLTVENPAGSKTVSKVLV	Db	11830	EIONYLEKETKRMVMSTYSATVLTPTGTTVTRLIEGNEYIFRVAENKIGTGPTTESKP
QY	98	-----97	QY	102	-----101
Db	10810	LDKPGPRDLEYSERKDSCLYLTWKEPLDGGSVITNVYVVERRDVASAQWSPLSATSKK	Db	11890	VIATKYDKRPPDPPEVTKVSKEEMTVVWNPPEYDGGKSITGYFLEKKEKHSTRWVPV
QY	98	-----97	QY	102	-----101
Db	10870	SHFAKHLNEGQVLFRAAENQYGRPFVETPKPIKALDPLHPPGPKDLHHVDVKTEV	Db	11950	KSAIPEREMKVQNLPLPOHEYQFRVKAENEIGIGBPSLPSRPVWAKDPIEPGPTNFRVV
QY	98	-----97	QY	102	-----101
Db	10930	SLVWPKPRDGGSPITGYLYVEYQEGTQDWIKFKVTNLECVVTGLOQKTYRFRVKAEN	Db	12010	DTTKHSITLGMGKPYDGGAPIIGYVVMRPKIADASPDGEMKRCNAAAQLVRKEFTVTS
QY	98	-----97	QY	102	-----101
Db	10990	IYVGLGLPDTTPIECOEKLVPSVELDVKLEGLVVKAGTTVRFPALIRGVPVPTAKWTT	Db	12070	LDENQYEFRCVCAQNVGIGRPAELKEAIKPKLELPEPEILDASMRKLIVIRACPIRL
QY	98	-----97	QY	102	-----101
Db	11050	DGSEIKTDEHYTVETDNFSSVLTIKNCLRRDTGEYQITVSNAGSKTVAVHLTVLDVGP	Db	12130	FAIVRGPRAPKVTRKVGIDNVVRKGQVLDVTMAFLVIPNSTRDSDGKYSLTLVNPA
QY	98	-----97	QY	102	-----101
Db	11110	PTGPNILVTPHEMTISWOPPKDDGGSPVINYIVEKQTRKDTWGVSSGSKTKLKIP	Db	12190	KAVFVNRVLDTPGPVSDLVKSDVTKTSCHVSWAPPENDGGSQVTHYIVEKREADRTWS
QY	98	-----97	QY	102	-----101
Db	11170	HLOKGEYVFRVAENKIGVGPLDSTPTVAKHKFSPSPGKPVVTDITENAAVSWTL	Db	12250	TVTPEVKTSFHVTLNPGNEYFRVTVAVNEYGCVPTDVPKVLASDPLSEDPKRLK
QY	98	-----97	QY	102	-----101
Db	11230	PKSDGGSPITGYMERREVTGKWRVNKTPIALDKFRVTGLYEGNTYEFVRFAENLAGLS	Db	12310	ATEMTKNSATLAWLPPLRDGGAKIDGYIISYREBEQPADRWTEYSVVKDLSLVVTGLKEG
QY	98	-----101	QY	102	-----101
Db	11290	KPSFSDPIKACRPKPPGPPINPKLKOKSRETADLVWTKPLSDGGSPILGYVVECKPG	Db	12370	KKYKFRVAARNAVSVSLPREAGVYEAKEQLLPKPKILMPEQITI KAGKKLR ¹ EAHVYKPK
QY	102	-----101	QY	102	-----101
Db	11350	TAQWNRINKDELIRQCAFVPGVGLIEGNEYFRRIKAANIVGEGEPRELAESVIAKDLHPP	Db	12430	HPTCKWKGDEVEVTS ¹ SHLAVHKADSSILIIKOVTRKDSGYYSILTAENSSGTDQKIKV
QY	102	-----101	QY	102	-----101
Db	11410	EVELDVTCDVITVRVGTIRILARVGRPEPDI ¹ TWTKEGKVLVREKRVLDI ¹ QDLPRVEL	Db	12490	VVMDAPGPPPPFDISDIDADACSLSWHIPLEDDGSGNITNYIVEKCDVSRGDWVTALASV
QY	102	-----101	QY	102	-----101
Db	11470	QIKA ¹ VRADHGKVIISAKNSSGHAQSGAIVNVLD ¹ RPGPCQNLKVTNVTKENC ¹ TSI ¹ WENPL	Db	12550	TKTSCRVGKLI ¹ PGQEYIFRVAENRFGISEPLTSPKMVAQFPFGVPSEPKNARVTKVND
QY	102	-----101	QY	102	-----101
Db	11530	DNGSEITNFIVEYRKNQKMSIVASDVTKRLIKANLLANNEYFRVCAENKVGVGPTI	Db	12610	CIFVANDRPDSGGSPII ¹ GYLIERKERN ¹ SLLVKANDTLVRSTEYPCAGLVEGLEYSFRI
QY	102	-----101	QY	102	-----101
Db	11590	ETKTPILAINPIDRGPENLHIADKGT ¹ FVYLKWR ¹ RPDYDGGSPNLSYHVERRLKGSDD	Db	12670	YALNKA ¹ SSPPSKPT ¹ EYVTARMPVDP ¹ PGKPEVID ¹ VTKSTVSLI ¹ WARPKHGGSKIIGFV

```
Db 12730 EACKLPDGKWRNCNTAPHQIQOEYATATGLEEKAQYQFRAIARTAVNISPPSESDPVTI 12789
Qy 102 ----- 101
Db 12790 LAENVPPRIDLSVAMKSLITVKGAGTNVCLDATVFGKPMPTVSMKKDGTLLKPAEGIKWAM 12849
Qy 102 ----- 101
Db 12850 ORNLICTLELFSVNRKSDGYTITAENSNGSKSATIKLVLDKPGPPASVKINKMYSDRAM 12909
Qy 102 ----- 101
Db 12910 LSWEPPLDGGSEITNIVDKRETSRNWAQVSATVPIITSCSVEKLEHGYQFICAEN 12969
Qy 102 ----- 101
Db 12970 KYGVGDPVFTEPAIAKPNYPDPGCRDPEVISNITKDHTVSWKPPADDDGGSPITGYLLEK 13029
Qy 102 ----- 101
Db 13030 RETQAVNWKVNRKPIIERTLKATGLOEGTEYEFRTVAINKAGPKPSDASKAAAYARDPQ 13089
Qy 102 ----- 101
Db 13090 YPPAPPAPKVYDTRRSVLSWGPAYDGGSPIIIGYLVKVRADSDNVVRCNLPQNLQK 13149
Qy 102 ----- 101
Db 13150 TRFEVTGLMEDTOYQFVYAVNKIGYSDPSDVPDKHYPKDILIPPEGEHADLRKTLILR 13209
Qy 102 ----- 101
Db 13210 AGVTMLRYVPVKGRPPPKITWSKPNVNLDRIGLDIKSTDFTFLRCENVNKYDAGKYL 13269
Qy 102 ----- 101
Db 13270 TLENSCGKKEYTIVVKVLDTGPPPINVTVKISKDSAYVTWEPPIIDGGSPINVVQKR 13329
Qy 102 ----- 101
Db 13330 DAERKSWSTVTTECKTSFRVPLNEBGKSYFPRVPAENEYIGDPGETRADVAKASQTPGP 13389
Qy 102 ----- 101
Db 13390 VVDLKVRSVKSSCSIGWKPHSDGGSRIIGYVVDFTLEENKQWRVMSLSLOYSAKDLT 13449
Qy 102 ----- 101
Db 13450 EGKEYTFRVSAENENGEGTPSEITVVARDVVAPDLDLGLPDLCLYLAKENSFRLKIPI 13509
Qy 102 ----- 101
Db 13510 KGKPAPSVKKGEDPLATDTRVSVSSAVNTLIVDCQKSDAGKYTITLKNVAGTKEG 13569
Qy 102 ----- 101
Db 13570 TISIKVKGPGIPTGPIKPFDEVTAEAMTLKMAPPKDDGGSEITNYILEKRDSVNNKWVTC 13629
Qy 102 ----- 101
Db 13630 ASAVOKTTFRTVRLHEGMEYTFRVSANKYGVGEGLKSEPIVARHPFDVPDAPPPNIVD 13689
Qy 102 ----- 101
Db 13690 VRHDSVSLTWDPKTKTGGSPITGYHLEFPKERNLSLLWKRAKNTPIRMDFKVTGLTEGLEY 13749
Qy 102 ----- 101
Db 13750 EFRVMAINAGVKPSLPSPEPVVALDPIDPGKPEVINITRNSVTLIWTEPKYDGHKLT 13809
Qy 102 ----- 101
```

```
Db 13810 GYIVBEKRDLPKSMKANHNVNPECAFTVTLVEGGKYEFRIRAKVTAGAISAPSESTET 13869
Qy 102 ----- 101
Db 13870 IICKOEYEAPTIVLOPTTIKDGTLTIKAGDTIVLNAISILCKPLPKSSWSKAGKDIRPSDIT 13929
Qy 102 ----- 101
Db 13930 QITSTPTSSMLTIKVATRKDAGEYITATNPFGTKVEHVKVTVLDVPGPPGPPVEISNVSA 13989
Qy 102 ----- 101
Db 13990 ERATLTWTPPLEDGGSPISKYILEKRETSRLLTWTVVSEDIQSCRHVATKLIOGNEYIFRV 14049
Qy 102 ----- 101
Db 14050 SAVNHYGKEPVQSEPVKWDPRFGPPGPEKPEVSNVTKNATATVSHKRPVDDGGSEITGY 14109
Qy 102 ----- 101
Db 14110 HVERREKSLRMVRAIKTPVSDLRCKVTGLQEGSTYEFVSAENRAGIGPPSEASDVLML 14169
Qy 102 ----- 101
Db 14170 KDAAYPPGPPSNPHVDTTTKSASLAWKPHYDGGLEITGYVVEHOKVQDEAMIKDOTTGT 14229
Qy 102 -LVITRF----- 107
Db 14230 ALRITQFVVVDLOTKEKYNFRISAINDAGVGEPAVIPDVEIVEREMAPDFELDAELRRTL 14289
Qy 108 ----- 107
Db 14290 VVRAGLSIRIFVPIKGRPAPEVTWTKDNINLKNRANIENTESFTLLIIPECNRYDTGKFV 14349
Qy 108 ----- 107
Db 14350 MTIENPAGKSGFVNVRLDRTARSPQLRPTDITKDSVTLHMDPLIDGSRITNYIVEK 14409
Qy 108 ----- 107
Db 14410 REATRKSYSTATTCKCHKCTYKVTGLSEGCEYFPRVMAENEYIGETPETTEPVKASEAPS 14469
Qy 108 ----- 107
Db 14470 PPSLNMIDITKSTVSLAWPKPKHGGSKITGYVIEAQRKGSQWTHITTVKGLECVVRN 14529
Qy 108 ----- 107
Db 14530 LTEGEYTFQWMAVNSAGRSAPRESRPVIVKEQTMPLPELDLRLGIYOKLVIAKAGDNIKVE 14589
Qy 108 ----- 107
Db 14590 IPVLRPKPTVTKGDOILLKQTORVNFETTATSTILNINECVRSDSGPYPLTARNIVGE 14649
Qy 108 ----- 107
Db 14650 VGDVITIQVHDIIPGPTGPIKFDEVSSDFVTFSDPPENDGGVPISNYVVMROTDSWTW 14709
Qy 108 ----- 107
Db 14710 VELATTVIRTTYKATRLTTGLEQYQFVKAQNRVYGVGPGITSAMIVANYPFKVPGPCTPQ 14769
Qy 108 ----- 107
Db 14770 VTAVTKDSMTISWHEPLSDGSPILGVYVERKERNGLWQTSKALVPGNIFKSSGLTDG 14829
Qy 108 ----- 107
Db 14830 IAYEFVRIAENMAGSKSPSEPMLALDPIPPGKPVPLNITRHTVTLKWAKEYTGGF 14889
Qy 108 ----- 107
Db 14890 KITSYIVVEKRDLPNGRWLKNFNSILENFTVSLGTEDAAVEFVRIAKNAAGAI5PSPSEP 14949
```

```
Qy 108 ----- 107
Db 14950 SDAITCRDDVEAPKIKVDYKFDTVILKAGEAFRLEADVSRPPTMEWSKDGELEGT 15009
Qy 108 ----- 107
Db 15010 KLEIKIADFSTNLNKNKSTRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPEGLAVTE 15069
Qy 108 ----- 107
Db 15070 VTSEKVLSPFPPLDDGGAKIDHIYIVQKRETSRLAWTNVASEVQVTKLVTKLLKGNEYI 15129
Qy 108 ----- 107
Db 15130 FRMAVNKYGVGELESEPLAVNPGPPDKNPEVTITKDSMVVCHGHPDSDGSGSI 15189
Qy 108 ----- 107
Db 15190 INVIVERRDKAGORWIKCNKKTLDLRYKVSGLTEGHEYEFRIMAENAAGISAPSPSPF 15249
Qy 108 ----- 107
Db 15250 YKACDVFVKPGPPGNPRVLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEEDEWQIVT 15309
Qy 108 ----- 107
Db 15310 PPAGLKATSYITGLTENQEYKIRIYAMNSEGLGEPALVPGTPKAEDRMLPPEIELDADL 15369
Qy 108 ----- 107
Db 15370 RKVVTIRACCTLRLFPVPIKGRDPPEVKWARDHGESLDKASIESAGSYTLIIVGNVNRFS 15429
Qy 108 ----- 107
Db 15430 GXYILTVNSSGSAFVNVNVLDTGPPQDLKVKEVTKTSVTLTWDPPLDGGSKIKNY 15489
Qy 108 ----- 107
Db 15490 IVEKRESTRKAYSTVATNCHKTSMWKVDQLEGCSYFVRVLAENEYIGLPAETAESVKAS 15549
Qy 108 ----- 107
Db 15550 ERPLPGKITLMDVTRNSVSLWEKPEHDGGSRIILYIVEMOTKSGDKWATCATVKVTEA 15609
Qy 108 ----- 107
Db 15610 TITGLIOGEEYSFRSAQNEKGISDPRQLSVPIAKDLVIPPAFKLLFNTFTVLAGEDLK 15669
Qy 108 ----- 107
Db 15670 VDVPPFIGRPTPAVTHKDNVPLKQTTTRVNAESTENNSSLTIKDACREDVGHVVKLTNSA 15729
Qy 108 ----- 107
Db 15730 GEAIETNLVILDKPGPTGPVKRWDEVTADSIITLSWGPVKPGYDSSINNYIVEKRDTSST 15789
Qy 108 ----- 107
Db 15790 TWOIVSATVARTTIKACRLKTGCEYQFRIAAENRYGKSTVYNSEPTVAQYPFVKVPGPGT 15849
Qy 108 ----- 107
Db 15850 PVTLSRDSMEVQWNEPISDGGSRVIGYHLERKERNISILWVKLNKTIPTQTKFKTTGLE 15909
Qy 108 ----- 107
Db 15910 EGVEYEFVRVSAENIVGIGKPSKSECVYARDPCDPPRPEAIIIVTRNSVTLOWKKPTYDG 15969
Qy 108 ----- 107
Db 15970 GSKITGYVEKKELPEGRWMAKSFNTIIDTHFEVTGLVEDHRYEFVRVIARNAAGVFSEPS 16029
```

```
Qy 108 ----- 107
Db 16030 ESTGAITARDEVDPRISMDFKYKDTIVVHAGESFKVDADIYGKPIPTIOWIKGDQLSN 16089
Qy 108 ----- 107
Db 16090 TARLEIKSTDFATSLSVKDAVRVDSGNILKAKNVAGERSVTNVKVLDRPGPEGPVVI 16149
Qy 108 ----- 107
Db 16150 SGVTAECTLAWKPLQDGGSDIINYIVERRETSRLVMTVVDANVQTLSCVKTKLLEGNE 16209
Qy 108 ----- 107
Db 16210 YTFRIMAVNKYGVGELESEPFVAKNPVVPDAPKAEVTTVTKDSMIVVWERPASOGGS 16269
Qy 108 ----- 107
Db 16270 EILGVLEKROKEGIRWTRCHKRLIGELRLAVTGLIENHDYEFVRVSAENAGLSEPPSPS 16329
Qy 108 ----- 107
Db 16330 AYQACDPIYKPGPPNPKVIDITRSSVFLSWSKPIYDGGCEIOGYIVEKCDVNVGWTM 16389
Qy 108 -----DL 109
Db 16390 CTPPTGINKTNIEVEKLEKHEYNFRICAINKAGVGEHADVPGPPIIVEEKLEAPDIDL 16449
Qy 110 EL----- 111
Db 16450 ELRKIINRAGGSLRLFPVPIKGRPTPEVKWKVDGEIRDAIIVTSSFTSLVLDNVNRY 16509
Qy 112 ----- 111
Db 16510 DSGKYTLTLENSGTSKSAFVTVRVLDTSPPPVNLKVTBITKDSVSIWEPPLDGGSKIK 16569
Qy 112 ----- 111
Db 16570 NYIVEKREATRKSAAVVTNCHKNWKIDQLEGCSYFVRVTAENEYIGILPAQTADPIK 16629
Qy 112 ----- 111
Db 16630 VAEVPOPPGKITVDVTRNSVSLSWTKPEHDGSKIIOYIVEMQAKHSEKSECARVKS 16689
Qy 112 ----- 111
Db 16690 QAVITNLTOGBEYLFRVVAVNNEKGRSDPRSLAVPIAKDLVIEPDVKPAPFSYSVQVGD 16749
Qy 112 ----- 111
Db 16750 LKMEVPISGRPKPTITWTKDGLPLKQTTTRINVTSDSLDTLTSIKETHKDDGGGYGITVAN 16809
Qy 112 ----- 111
Db 16810 VVGQXTASIEIVTLDKPDPKPGVKFDVDSAESITLSWNPPLYTGCCQITNVIQKRDIT 16869
Qy 112 ----- 111
Db 16870 TTVMDVVSATVARTTLKVKLTGTGYOFRIFAENRYGQSFALSDPIVAQYYPYKEPGPP 16929
Qy 112 ----- 111
Db 16930 GTPFATAISKDSMVJQWHEPVNNGSRVIGYHLERKERNISILWTKVNTIHDTOFKAON 16989
Qy 112 ----- 111
Db 16990 LEEGIEYEFVRVYAENIVGVKASKNSECVYARDPCDPPGTPEPIMVKRNEITLOWTKPVY 17049
Qy 112 ----- 111
Db 17050 DGGSMITGYIVEKROLPGRWMAKSFNTNVIETQFTVSGLTEDQRYEFVRVIKNAAGAIK 17109
Qy 112 ----- 111
```


Db	17110	PSDSTGPITAKDEVLPRI	SMOPKFRD	TI	VNAGETFRLEADVHGKPLPTIEWLRGDK	17169
Qy	112	-----	-----	-----	-----PDGNR	116
Db	17170	EESARCEIKNTDPKALL	IVKDAIRIDGGQYIL	RASN	VAGSKSPVNVKVLDRPGP	17229
Qy	117	QVRGVT	-----	QLGG	-----	126
Db	17230	QVTGVSEKCSLTWS	PLQDGGSDISHYV	VEKRETSRLAWTV	VASEVVTNSLKVTKL	17289
Qy	127	-----	-----	-----	-----	126
Db	17290	NEYVFRIMAVNKY	GVGBLESAPVLM	MKNPFVLP	GPCKSLEVTNIAKDSMTVC	17349
Qy	127	-----	-----	-----	-----	126
Db	17350	GSEIIGVIVEKRD	RGIRWIKCNKR	RITDLRLRV	TGLTEDEHEYFRVSA	17409
Qy	127	-----	-----	-----	-----	130
Db	17410	ATVYKACDPVK	PGPTNAHIVDTTK	NSITLAWGKPIY	DGSGEILGYVVEICKAD	17469
Qy	131	-----	-----	-----	-----	130
Db	17470	QIVTPOTGLRV	TRPEISKLTEHQ	EYKIRVCALNK	VGLGEATSVPGTVK	17529
Qy	131	-----	-----	-----	-----	133
Db	17530	DSELRKGI	VWAGGSARIH	IPFKRPMPEIT	WSEEGEFTDKVQIEK	17589
Qy	134	-----	-----	-----	-----	133
Db	17590	RNDAGKYL	KLENSGSKSA	FVTVKVLDT	PGPPQNLA	17649
Qy	134	-----	-----	-----	-----	133
Db	17650	VKNYVIDKRE	STRKAYANVSS	KSKTSFKVENL	TEGAIYYFRVMA	17709
Qy	134	-----	-----	-----	-----	133
Db	17710	VKAAPSP	PGKVLT	TVDSQTSAS	LMWEKPEH	17769
Qy	134	-----	-----	-----	-----	133
Db	17770	VCNAVVTGL	SSQEQYQFR	KAYNEKGS	DPVVLGVP	17829
Qy	134	-----	-----	-----	-----	133
Db	17830	EDLKIEIP	VIGRPRN	ISHWKDEPL	KQTRVNV	17889
Qy	134	-----	-----	-----	-----	133
Db	17890	TNSAGTAT	ENLSVIVLEK	PPVGPVR	DEVSADFV	17949
Qy	134	-----	-----	-----	-----	133
Db	17950	TTTTTHMV	SATVARTTI	KITKLTG	EYQFRIFA	18009
Qy	134	-----	-----	-----	-----	133
Db	18010	PGTPEVTS	ISKDQMLV	QWHEP	VNDGGTKI	18069
Qy	134	-----	-----	-----	-----	133
Db	18070	TGLDEGLE	YFKVSA	ENIVGICK	PSKVSCE	18129
Qy	134	-----	-----	-----	-----	133
Db	18130	AYDGGSKI	TYIV	VEKKDLP	GRMKAS	18189
Qy	134	-----	-----	-----	-----	133
Db	18190	SEPDSS	GAITAR	DEIDAP	NASLDPKY	18249
Qy	134	-----	-----	-----	-----	133
Db	18250	ELEETA	ARMEIK	STIQKTT	LVVKDC	18309
Qy	134	-----	-----	-----	-----	133
Db	18310	GPLKVT	GTAECY	LAWN	PLQDGGAN	18369
Qy	134	-----	-----	-----	-----	133
Db	18370	LPGNE	YIFR	MAV	NKYGIG	18429
Qy	134	-----	-----	-----	-----	133
Db	18430	DDGTE	IEGY	ILEK	RDKEG	18489
Qy	134	-----	-----	-----	-----	133
Db	18490	PSEPS	VYRAC	DALYP	PPGPP	18549
Qy	134	-----	-----	-----	-----	133
Db	18550	DEWTT	CTPT	PTGL	QKQFTV	18609
Qy	134	-----	-----	-----	-----	133
Db	18610	IELDAD	LRKV	VL	RASAT	18669
Qy	134	-----	-----	-----	-----	133
Db	18670	NVTR	PD	SGRY	NLTENN	18729
Qy	134	-----	-----	-----	-----	133
Db	18730	GAKIT	NVIV	EKRE	TRKAY	18789
Qy	134	-----	-----	-----	-----	133
Db	18790	TEPV	KVSE	PL	PPGR	18849
Qy	134	-----	-----	-----	-----	133
Db	18850	QVKT	EATIS	GL	TAGEE	18909
Qy	134	-----	-----	-----	-----	133
Db	18910	KARE	QKID	VFP	KGR	18969
Qy	134	-----	-----	-----	-----	133
Db	18970	LCV	NSAG	SIT	VP	19029
Qy	134	-----	-----	-----	-----	133
Db	19030	KETT	STTH	IV	SQAV	19089
Qy	134	-----	-----	-----	-----	133
Db	19090	PGP	PTK	VV	HATK	19149
Qy	134	-----	-----	-----	-----	133
Db	19150	KVSG	LDE	GL	MYE	19209
Qy	134	-----	-----	-----	-----	133
Db	19210	KPHY	DGGA	KIT	GI	19269
Qy	134	-----	-----	-----	-----	133
Db	19270	SVSE	PSE	ST	GP	19329
Qy	134	-----	-----	-----	-----	133

Qy	134	-----	133
Db	19330	GIEIERARTEIISTDNHTLLTVKDCIRRDTOGYVLTLLKNVAGTRSAVNCVKLOKPGPP	19389
Qy	134	-----	133
Db	19390	AGPLEINGLTAEKCSLSWRPOEDGGADIDYYHRKKRETSHLAWTICEGELQMTSCVKTK	19449
Qy	134	-----	133
Db	19450	LLKGNEYIFRVTVGNKYGVGEPLSAIKALDPFTVPSPPTSLEITSVTKESMTLCWSRP	19509
Qy	134	-----	133
Db	19510	ESDGGSEISGYIIERREKNSLRWRNKKRPVYDLRVKSTGLREGCEYEYRVYAENAAGLS	19569
Qy	134	-----	133
Db	19570	LPSETSPLIRAEPPVFLPSPSPKPIVDGKTTITIAWKPLFDGGAPITGTVVEYKSD	19629
Qy	134	-----	133
Db	19630	DTDWKTSIOSLRGTEYTIISGLTTGAEYVFRVKSVMKVGASDPSSDPQIAKEREPEPLF	19689
Qy	134	-----	133
Db	19690	DIDSEMRKTLIVKAGASFTWTVFRRGPVNVVLSKPPDTRLTRAVVDTDSRTSLTIEN	19749
Qy	134	-----	133
Db	19750	ANRDSKGYTLTIQNLVLSAASLTLVVKVLDTPGPPNTITVQDVTKESAVLSMDVPENDGG	19809
Qy	134	-----	133
Db	19810	APVKNVHIEKREASKAWSVTNNCNRLSYKVTNLOEGAIYYFRVSGENEFVGVI PAETK	19869
Qy	134	-----	133
Db	19870	EGVKITEKSPPEKJGVTISKDSVSLTWLKPHEHGGSRIVHYVVEALEKQGNWVCAV	19929
Qy	134	-----	133
Db	19930	AKSTHHVVSGLRESEYFFRVFAENQAGLSDPRELLLPVLIKEQLEPEIDMKNFPSTHTV	19989
Qy	134	-----	133
Db	19990	YVRAGSNLKVDDIPISGKPLPKVTLSDRGVPLKATMRENTITAENLTINLKESVTADAGR	20049
Qy	134	-----	133
Db	20050	YEITAANSSGTTKAFINIVVLDPRGPTGPVVISDITEESVTLKWEPKPYDGSQVTNYI	20109
Qy	134	-----	133
Db	20110	LLKRETSAVTVESATVARTMKVMKLTGTGEEYQPRIKAENRFGISDHIDSACVTVKLP	20169
Qy	134	-----	133
Db	20170	YTPGPPSTPWVNTVRESITVGHPEPVNSGSAVVGHYHLEMKDRNSILWQKANKLVIRT	20229
Qy	134	-----	133
Db	20230	THFKVTTISAGLIYEFRYAENAAGVKPSHPSEPVLAIDACEPPNRVTRIDISKNSVSL	20289
Qy	134	-----	133
Db	20290	SWQOPAFDGGSKITGYIVERDLDPGRWTKASFNTVTTQTTISGLTQNSQYERFVPARN	20349
Qy	134	-----	133
Db	20350	AVGSI SNPSEVVGPI TCIDSYGGPVIDLPLEYTEWVKYRAGTSVKLRAGISGKPAPTIEW	20409

Qy	134	-----	133
Db	20410	YKDDKELQTNALVCVENTTDLASILKOADRLNSGCYELKLRNAMASASATIRVQILDKP	20469
Qy	134	-----	137
Db	20470	GPPGGPIEFKVTATBKITLLWRPPADGGAKITHYIVREKRETSRVVMSVSEHLEECIIT	20529
Qy	138	-----	137
Db	20530	TTKIIKNEYIFRVRAVNNKYGIGEPLESDSVVAKNAFVTPGPPGIPVTKITKNSMTVVM	20589
Qy	138	-----	137
Db	20590	SRPIADGGDISGYFLEKRXKSLGWFKVLTETIRDTQKVTGLTENSQYQYRVCAVNA	20649
Qy	138	-----	137
Db	20650	QGGPFSEPFYKAADPIDPPGPPAKIRIADSTKSSITLGMSPKPVYDGGSAVTGVVVEIR	20709
Qy	138	-----	137
Db	20710	QGEEBEWTTVSTKGEVRTTEYVVSNLKPGVNYVFRVSAVNCAGQGEPIEMNEPVQAKDIL	20769
Qy	138	-----	137
Db	20770	EAPRIDLDVALRTSVIAKAGEDVQVLIPFKGRPPPTVTRKDEKMLGSDARYSIENTDSS	20829
Qy	138	-----	137
Db	20830	SLLTIPQVTRNDTGKIYIITENGVEPKSVSVKVLDTPAACQKLVKHVSRGTVTLLW	20889
Qy	138	-----	137
Db	20890	DPPLIDGGSPINYYIEKRDATKRTWSVSHKCSSTSFKLIDLSEKTPFFFRVLAENEIG	20949
Qy	138	-----	137
Db	20950	IGPCETTEPVKAAEVPAPIRDLNMDSTKTSVILSWTKPDPDGGSVITEYVVERKSGE	21009
Qy	138	-----	137
Db	21010	QWSHAGISKTEIEVSOLKEQSVLEFRFAKNEKGLSDPVTIGITVKELIITBEVDLS	21069
Qy	138	-----	137
Db	21070	DIPGAQVTVRIGHNVHLELPYKGPSPISWLKQGLPKESEFVRFSKTENKITLSIKVA	21129
Qy	138	-----	137
Db	21130	KXEHGKVTVILDNAVCRIVPITVITLGPSPKGPPIRFBIDIKADSIVLSWDVPEDNGG	21189
Qy	138	-----	137
Db	21190	GEITCYSIEKRETSOTNNWMCSSVARTTFKVPNLVKOAEYOFVRAENRYGVSQPLVSS	21249
Qy	138	-----	137
Db	21250	IIVAKHQFRI PGPPGPKPIYNTVSDGMSLTWDAPVYDGGSEVTGFHVEKKERNLSLWQKV	21309
Qy	138	-----	137
Db	21310	NTSPISGREYRATGLVEGLDYQFRVYAENSAGLSSPSPDKFTLAVSPVDPPTDYIDV	21369
Qy	138	-----	137
Db	21370	TRETITLKNPPLRDGSKIVGYSIEKROGNERWVRNFTDVSECOYTVTGLSPGDRYEF	21429
Qy	138	-----	137
Db	21430	RIIARNAVGTISPPSSSGIIMTRDENVPPIVEFGPEYFDGLIIKSGESIRIKALVQGRP	21489
Qy	138	-----	137

Db	21490	VPRVTFKDGVEIEKRMMEITNVLGSTSLFVRDATROHRCVYVTEAKNAGSAGAEIKV	21549
Qy	138	-----	137
Db	21550	KVQDTPCKVVGPIRFTNITGEKMTLMDAPLNDGCAPITHYIEKRETSRLAWALIEDKC	21609
Qy	138	-----	137
Db	21610	EAQSYTAIKLINGNEYQFRVSANVKFGVGRPLDSPVVAQIQYTVDPAGIPEPSNITGN	21669
Qy	138	-----	137
Db	21670	SITLTWARPESDGGSEIQOYLIERREKKSTRVVKVISKRPISETRPKVTGLTEGNEYEPH	21729
Qy	138	-----	137
Db	21730	VMAENAAVGVPASGISRLIKCREPNPPGPTTVVKVTDTSKTTVSLEWSKPVFDGMBII	21789
Qy	138	-----	137
Db	21790	GYIEMCKTDLGDMHKVNAEACVTRYTVTDLQAGEYKFRVSAINGAGKGDSCVETGTI	21849
Qy	138	-----	137
Db	21850	KAVDLTAPELDIDANFKQTHVVRAGASIRLFIAYQGRPTPTAVWSKPDNSLSURADIHT	21909
Qy	138	-----	137
Db	21910	TDSFSTLTVCNCRNDAGKTYLTVENNSGSKSIITFTVKVLDTPGPPGPIITPKDVTGRSAT	21969
Qy	138	-----	137
Db	21970	LMWDAPLDGGARIIHHVYVVEKREASRRSQVISEKTRQIIFKVDNLAEQVPIYFRVSANV	22029
Qy	138	-----	137
Db	22030	EYGVGEPEMPEPIVATEQPAPRRLDVVDTSKSSAVLAWLKPDDHGGSRITGYLLEMQR	22089
Qy	138	-----	137
Db	22090	KGSDLWEAGHTKQLTFTVERLVEKTEYEFVKAKNDAGYSEPREAFSSVIIKEPQIEPT	22149
Qy	138	-----	137
Db	22150	ADLTGITNQLITCKAGSPFTIDVPISGRPAKV/TWKLEEMRLKETDRVSIITTKDRTTLT	22209
Qy	138	-----	137
Db	22210	VKDSMRGDSGRYFLTLENTAGVKTFSVTVWVIGRPGVPTGPIEVSSVSAESCVLWSGEPK	22269
Qy	138	-----	137
Db	22270	DGGGTEITNIYIVEKRESGTTAQLNVSSVKRTOIKVTHLTKYMEYSFRVSSENRFGVSKP	22329
Qy	138	-----	137
Db	22330	LESAPIIAEHPPVPSAPTRPREVYHVSANAMSIIRWEEPHYDGGSKIIGYWVEKKERNITL	22389
Qy	138	-----	137
Db	22390	WYKENKVPCLCNYKVTVGLVEGLEYOFRTYALNAAGVSKASEASRPIMQNPVDAPGRPE	22449
Qy	138	-----	137
Db	22450	VTDVTRSTVLSIWSAPAYDGSKWGYIIERKPVSEVGDRWLKCNKYTIIVSDNFTVTAL	22509
Qy	138	-----	137
Db	22510	SEGDTYEFRLAKNAAGVTSKSESTGPVTCRDEVAPPKAELDARLHGLDLVTIRAGSDLV	22569
Qy	138	-----	137
Db	22570	LDAAVGGKPEPKI IWTGDKELDCEKVSLOYTGKRATAVIKFCDRSDSGKYTLTVKNAS	22629
Qy	138	-----	137
Db	22630	GTKAVSMVKVLDSPGCGKLTVSRVTQEKCTLANSLPOEDGGAETHYIVERRETSRLN	22689
Qy	138	-----	137
Db	22690	WVIVEGECPTLSYVVTRLIKNNEYIFRVRANVKYGPVPESEPIVARNSTIIPSPGP	22749
Qy	138	-----	137
Db	22750	EEVGTKEHII IOWTKPESDGGNEISNYLVDKREKSLRWTRVKNKYVVDYDRLKVTSLM	22809
Qy	138	-----	137
Db	22810	EGCDYQFRVTAVNAAGNSBPSERSNFISCREPSYTPGPPSAPRVVDTTKHSISLAWTKPM	22869
Qy	138	-----	137
Db	22870	YDGGTDIVGVLEMOBKTDQWYRVHTNATIRNTEFTVPDLKMGQKYSFRVAANVKWMS	22929
Qy	138	-----	137
Db	22930	EYSESAEIEPVERIEIPOLELADLLKKTVTIRAGASLRLMYSVSGRPPPVITMSKQID	22989
Qy	138	-----	137
Db	22990	LASRAIIDTTESYSLIIVDKVNRDAGKVTIEAENQSGKKSATVLVKVYDTPGCPSPVKV	23049
Qy	138	-----	137
Db	23050	KEVSRDSVTITWBIPTIDGGAPINNYIVEKREAAAFKTVTTKCKSTLYRISGLVEGTM	23109
Qy	138	-----	137
Db	23110	HYFRVLPENIYGEGECETSDAVLVSEVPLVPAKLEVDVTKSTVTLAWEKPLYDGSRL	23169
Qy	138	-----	137
Db	23170	TGYVLEACKAGTERMKVVTLKPTVLEHTVTSLEGEQVLFIRAQNEKGVSEPRETVTA	23229
Qy	138	-----	137
Db	23230	VTQDLRVLPTIDLSTMPQKTIHVPAGRPVELVIAGRPAAASWFFAGSKLRESERV	23289
Qy	138	-----	141
Db	23290	WETHKVAKLTIRETITRTDGEYVLEKNVTGTTSETIKVIIIDKPGPTGPIKIDEIDA	23349
Qy	142	-----	141
Db	23350	TSITISWEPPELDGAPLSGVVVEQDAHRPGWLPVSESVTRSTFKFTRLTEGNEYVFRV	23409
Qy	142	-----	141
Db	23410	AATNRFIGSYLOSEVIECRSSIRIPGPETLOIFDVSRDGMTLTWYPEDDGGSQVTGY	23469
Qy	142	-----	141
Db	23470	IVERKEVRADRVRVKNVPTWTRYSTGLTEGLEYEHRVTAINARGSKPSPKPIVA	23529
Qy	142	-----	141
Db	23530	MDPIAPGKPNQPRVTDTRTSVSLAWSVPEDEGSKVTGYIEMOKVQDHEMTKCNTP	23589
Qy	142	-----	141
Db	23590	TKIREYTLHLPOGAEYRFRVLACNAGGPEPAEVPQTKVTEMLEYPDYELDERYQEG	23649
Qy	142	-----	141
Db	23650	FVROGGVIRLTIPIKGRPFPICKWKTEGQDISKRAMIATSETHTELVIKEADRGDSCTYD	23709

QY 142 ----- 141
 Db 23710 LVLENCKGKANYIKVRVIGSPNSPEGPLEYDDIQVRVSVRPPADGGADILGYILE 23769
 QY 142 ----- 141
 Db 23770 RREVPKAAWYIDSRVRGTSLVVKGLKENVEYHFRVSAENQFGISKPLKSEEPVTPKTEL 23829
 QY 142 ----- 141
 Db 23830 NPPEPPSPPEVLDTKSSVLSWSRPKODGSRVTGYIETKSTDKVVRHNKTOITT 23889
 QY 142 ----- 141
 Db 23890 TMYTVGLVDAEYQFRITANDVGLSETSPASEPVVCKDPFDPKPSQGELEILSISKDS 23949
 QY 142 ----- 141
 Db 23950 VTLQWEKPECDGCKEILGYWVEYRQSDSAWKSNKERIKDKQFTIGGLLEATEYEFVRF 24009
 QY 142 ----- 141
 Db 24010 AENETGLSRPRRTAMSIKTKLTSGEAPGIRKEMKDVTTKLGEAAQLSCQIVGRPLDIKW 24069
 QY 142 ----- 141
 Db 24070 YRFGKELIQSRKYKMSDGRHTLTVMTERQEDGVYTCIATNEVEGETSSKLLQATP 24129
 QY 142 ----- 141
 Db 24130 QHPGYPLKEKYGAVGSTLRLHWYIGRPVPMATWPHGQKLLQNSENTIENTEHTL 24189
 QY 142 ----- 141
 Db 24190 VMKNVORKTHAGRYKVQLSNVFGTVDAILDVEIQDKPKPTGPIVIEALLKNSAVISWKP 24249
 QY 142 ----- 141
 Db 24250 PADGGSWITNVVVEKCEAKEGAELVSSAISVTTCRIVNLNENAGYVFRVSAQNTFGI 24309
 QY 142 ----- 141
 Db 24310 SDPLEVSSVVIKSPPEKPGAPKPTITAVTKDSCVVAMKPPASDGGAKIRNVYLEKREK 24369
 QY 142 ----- 141
 Db 24370 KONKWSVTTTEIRETVFSVKNLIEGLEEFVRVKCENLGGESEWSEISEPITPKSDVPIQ 24429
 QY 142 ----- 141
 Db 24430 APHPKEELNLRVYQSNATLVCKVCHPKPIVKNYRQGEIITADGLKYRIQEFKGYHQ 24489
 QY 142 ----- 141
 Db 24490 LIASVTDDATVYQVRATNGGVSCTASLEVEVPKIHLPKTLQMGAVHALRGEWS 24549
 QY 142 ----- 141
 Db 24550 IKIPFGKPDVITWQKQDLIDNNGHYQVITRSFTSLVPFNGVERKDGAFYVVCANR 24609
 QY 142 ----- 141
 Db 24610 FGIDQKTVELDVADPPDPGRGVKVSASRDSVNLTWTEPDSGSKITNIVEKATTAE 24669
 QY 142 ----- 141
 Db 24670 RWRVQARETRYVINLFGKTSYQFRVIAENKFGLSKPSPEPTITKEDKTRAMNDE 24729
 QY 142 ----- 141
 Db 24730 EVDETREVSMTKASHSSTKELYEKYEMIAEDLGRGEFGIVHRCVETSSKKTVMKVFVKVG 24789

QY 142 ----- 141
 Db 24790 TDQVLVKEISILNIARHNILHLHESFESMEELVMIFEFGSLDIFRINTSAFELNER 24849
 QY 142 ----- 144
 Db 24850 EIVSVHVOCEALQPLSHSHNIGHFDIRPENIYQTRRSSTIKIIEFGOARQLKPGDNFRL 24909
 QY 145 ----- 144
 Db 24910 LFTAPEYVAPEVHQHDVVSTATDMWSLGLTVVYLLSGINPFLAETNOOIENIMNAEYTF 24969
 QY 145 ----- 144
 Db 24970 DEEAFKEISIEAMDFVDRLLVKERKSRMTASEALQHPMLKOKIERVSTKVIRTLKHRRY 25029
 QY 145 ----- 144
 Db 25030 HTLIKKDLNMVVSAAIRISCGGAIRSQGVSAKVVASIEIGPVSQIMHVAEGEGHVK 25089
 QY 145 ----- 144
 Db 25090 YVCKIENVYDQSTQVTMYEGVRQLENSEKYEITYEDGVAIYVKDITKLDGTYRCKVND 25149
 QY 145 ----- 144
 Db 25150 YGEDSSYAELFVKGVREVVYCYCRRTMKIKRRTDMRLERPEFTLPLYNKTAIVGEN 25209
 QY 145 ---GVTI 148
 Db 25210 VRFQVTI 25216
 RESULT 2
 S41309
 Cyclosporin synthetase - cyclosporin fungus
 C:Species: Tolypocladium inflatum (cyclosporin fungus)
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 03-Nov-2000
 C:Accession: S45487; S41309
 R:Weber, G.; Schoergerdorfer, K.; Schneider-Scherzer, E.; Leitner, E.
 Curr. Genet. 26, 120-125, 1994
 A:Title: The peptide synthetase catalyzing cyclosporine production in Tolypocladium n
 A:Reference number: S45487
 A:Accession: S45487
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-15281 <WEB>
 A:Cross-references: EMBL:Z28383; NID:G440168; PIDN:CAA82227.1; PID:G440169
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993
 A:Note: only a part of the translation is shown
 A:Note: the source is designated as Tolypocladium inflatum
 C:Genetic code: SGC3
 A:Genetic code: SGC3
 C:Superfamily: cyclosporin fungus cyclosporin synthetase; acetate-CoA ligase homology
 C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
 F:532-1008/Domain: acetate-CoA ligase homology <ACLI>
 F:1029-1096/Domain: acyl carrier protein homology <ACP1>
 F:1618-2069/Domain: acetate-CoA ligase homology <ACLI2>
 F:2527-2594/Domain: acyl carrier protein homology <ACP2>
 F:3115-3553/Domain: acetate-CoA ligase homology <ACP3>
 F:4014-4081/Domain: acyl carrier protein homology <ACP4>
 F:4601-5057/Domain: acetate-CoA ligase homology <ACLI4>
 F:5506-5573/Domain: acyl carrier protein homology <ACP4>
 F:6094-6546/Domain: acetate-CoA ligase homology <ACLI5>
 F:7003-7070/Domain: acyl carrier protein homology <ACP5>
 F:7591-8042/Domain: acetate-CoA ligase homology <ACP6>
 F:8063-8130/Domain: acyl carrier protein homology <ACP6>
 F:8652-9157/Domain: acetate-CoA ligase homology <ACP7>
 F:9558-9625/Domain: acyl carrier protein homology <ACP7>
 F:10146-10586/Domain: acetate-CoA ligase homology <ACLI8>
 F:11055-11122/Domain: acyl carrier protein homology <ACP8>
 F:11635-12106/Domain: acetate-CoA ligase homology <ACLI9>
 F:12127-12194/Domain: acyl carrier protein homology <ACP9>

F:12715-13159/Domain: acetate-CoA ligase homology <AL10>
F:13623-13690/Domain: acyl carrier protein homology <AC10>
F:14213-14676/Domain: acetate-CoA ligase homology <AL11>
F:14698-14765/Domain: acyl carrier protein homology <AC11>
F:1060.2558.4045.5537.7034.8094.9589.11086.12158.13654.14729/Binding site: phosphopantec

Query Match 34.28; Score 264; DB 2; Length 15281;

Best Local Similarity 0.84; Pred. No. 17;

Matches 102; Conservative 24; Mismatches 21; Indels 12475; Gaps 32;

Qy 1 AAGGIL----- 6
||| :|
Db 2098 AAGHVLEIGTGTGVLFNLGAGLKS YIGLEPSQSAQVFNKAAQTFGLEKQAQVHGT 2157
Qy 7 ----- 6
Db 2158 AMDTGRSALSPDLIVNSVAQYFPSREYLAEVVEALVRIPGVRRIFFGDMRTYATHKDF 2217
Qy 7 -----HLELL----- 11
||| :|
Db 2218 LVARAVHTNGSKVTRSKVQVEARLEELLEVDPAPFTSLKESLEEIEHVEILPKNM 2277
Qy 12 ----- 11
Db 2278 KVNNELSSRYGAVLHIRNHNQNSRSHKINAESWIDFASSQMDRQGLARLLKENKDAE 2337
Qy 12 ----- 11
Db 2338 SIAVENIPYSKTI VERHIAKSLADHDGDDTHSSIDGVAMISAAREKASQCPSLDVHDLV 2397
Qy 12 ----- 11
Db 2398 QLAEDAGFRVEVSUARQSQNGALDVFFHHFOPTENESRALVDFTDYKGOQARSLTNRP 2457
Qy 12 ----- 11
Db 2458 LQVESRRIEAQVREQLVLLPAYMPIPARIVVLQNPLNTSGKVDKELTLRAKVTAART 2517
Qy 12 ----- 11
Db 2518 PSSELVAPRDSIEAIIKCEKFDVLGVEVGITDNFFNVGHSLLATKLAARLSRQLNAQIA 2577
Qy 12 ----- 11
Db 2578 VKDIFDRPVIADLAATIOQDTTEHPILPTSYTGPVEQSAQGRFLWFLDQLNVGATWYLM 2637
Qy 12 ----- 11
Db 2638 PFAVRLGPLVVSALAAALLAEERHETLRTTFIEQEGIGMQVIHPFAPKELRVIDVSGE 2697
Qy 12 ----- 11
Db 2698 EESTIQILEKEQTFFNLASEPGFRLALLKTGEDEHILSTVMHHAISDGWSVDIFQOEI 2757
Qy 12 ----- 11
Db 2758 GOFYSAILRGHDP LAQIAPLSIQYRDFATWQRQIFQVAEHRRLQAYTKQLADNKP AELL 2817
Qy 12 ----- 11
Db 2818 TDFKRPMLSGRAEIPVVDGLIYEKLQDFCRIRQVTAFTVLLAAFRAAHYRMTGTEDA 2877
Qy 12 ----- 11
Db 2878 TIGTPIANRNRELEGLIGFFVNTQCMRTVDVDSFETLVHVRRETTLAHANODVPFE 2937
Qy 12 ----- 11
Db 2938 QIVSNILPGSSDTSRNPLVQLMFAHLSQQNLKVRLEGIEEIIISIAETTRFDEIHFHYQ 2997
Qy 12 ----- 11
Db 2998 EAERLNGSIVYAADLFVPETIQSVITIFQGILOKGLGEPMFVSMALDGGLESLSRSTGL 3057

Qy 12 -----VAVGPDV----- 18
||| :|
Db 3058 LHPQQTDPDASVVOIFKQAVAVNPDVAVRDESTRLSVADLRKSDQVACHLSRRGIA 3117
Qy 19 ----- 18
Db 3118 PETFVAILAPRSCETIVAILGVLKANLAYLPLDVNVPASRLEAILSEVSGSMLVGAET 3177
Qy 19 ----- 18
Db 3178 PIPEGMAEATIRITEILADAKTDDINGLAASOPTAASLAYVIFTSGSTGRPKGVMEHR 3237
Qy 19 ----- 18
Db 3238 CIVRLTKQTNITSKLPESFHMHI SNLAPDASVMEVFTLLNGGTLVCIDYFTLLESTAL 3297
Qy 19 ----- 18
Db 3298 EKVFQDQVNVALLPPALLKQCLDNSPALVKLTSLVLYIGDRLDASDAKAKARGLVOTQAF 3357
Qy 19 ----- 18
Db 3358 NAYGPTENTVMSTIYPIAEDPFINGVPIGHAVNSGAFVMDNQOITPPGAMGELIVTGD 3417
Qy 19 ----- 18
Db 3418 GLARGYTTSLNTGRFINVDIDGEOVRAYRTGDRVVRPKDLOIEFFGRIDHVKIRGHR 3477
Qy 19 ----- 18
Db 3478 IEPAEVEYALLSHDLVTDAAVVVTHSOENODLEMVGFVAARVADVREDESSNOVQEWTHF 3537
Qy 19 ----- 18
Db 3538 DSIAYADITTDOOSLGRDFMSWTSMYDGLIKKSOMQEWLDDTMRSLLDSQPPGHVLEV 3597
Qy 19 ----- 18
Db 3598 GTGTGMVLFNLGREGGLOS YVGLPEPSATAFVFNKAAKSPGLEDRIRVEVGATDIDL 3657
Qy 19 ----- 18
Db 3658 GDDLHAGLVVNSVAQYFPSQDYLAQLVRLTKVPGVERIFFGDMRSHAINRDFLVARAV 3717
Qy 19 ----- 18
Db 3718 HALGDKATKAEIOREVVRMESEDELLVDPAFTSLTTQVENIKHVEILPKRMATNELS 3777
Qy 19 ----- 18
Db 3778 SYRYAAVLHVNDLAKPAHKVSPGANVDPAAATKMDRDALIRLLRGTKISDHIAIANIPNSK 3837
Qy 19 ----- 18
Db 3838 TIVERTICESVYDIGGDAKSDNRVSWLSAARSNAVKVASLSAIDLVDIAQEAQFRVEIS 3897
Qy 19 ----- 18
Db 3898 CAROMSONGALDAVFFHHLGSPSSSHVLIDFLTDHQRPEALTNHPHRAQSRVERQI 3957
Qy 19 ----- 18
Db 3958 RERLQTLTPAYMPIAQIMVLDKLPLNANGKVDKQLTORAQVTPKAKOVSAVPAPRTEIE 4017
Qy 19 ----- 18
Db 4018 RVLCOEFSVDLGVDIGIMENFFDLGHSLSMATKLAARISRRLETHSVSKEIFDHPRVCDL 4077
Qy 19 ----- 18
Db 4078 VLIVOQGSAPHDPVISTKYTGVPDQSPQAFQRLWFLDQNFATWYLMPLAVLRGAMNVH 4137

QY 19 ----- 18
Db 4138 ALTAALLERHLLRTTFYEQNGVGMKNPVVTTETLRIIDLSNGDGYLPTLKKEQT 4197
QY 19 ----- 18
Db 4198 APFHLETEGMRVALLRGLPGDYLILSVVHHIISDQWSVDVLFQELGQFYSTAVKGHDPL 4257
QY 19 ----- 18
Db 4258 SOTTPLPIHYRDFALWOKPTOESEHEROLQYWEQVDSAPAELLTDLPRPSILSGQAG 4317
QY 19 ----- 18
Db 4318 EMSVTIEGALYKNLEBFRCVHRVTSFVLLAALRAAHYRLTGSSEDATIGTPIANRNPRL 4377
QY 19 ----- 18
Db 4378 EQIGFFVNTQIRITVNEDETFSILVQVRSTATAAFAHQDVPEKIVSTLLPGSRDAS 4437
QY 19 ----- 18
Db 4438 RNPLVOLMPFAVHSQKNLGELKENAHSEVVPTEITTRFDLEFHLFOODDKLEGSILYSTD 4497
QY 19 ----- 18
Db 4498 LFEAVSQSLLSVFQELIRRLGNGDPVPISTLPLODGIYDLQROGLLDVQKTEYPRDSSV 4557
QY 19 ----- 18
Db 4558 VDFHEQVSINPDSIALIHGSEKLSYAQDRESDRVARWLRHRSFSSDTLIAVLAPRSCE 4617
QY 19 ----- 18
Db 4618 TIAFLGILKANLAYLPLDVKAPARIDAIVSSLPKNKILILGANVTPPKLOEAAIDFVP 4677
QY 19 ----- 18
Db 4678 IRDTFTLLDGTLODGTIERPSAQSLAYAMFTSGTRPKGMVQHRNIVRLVKNSVV 4737
QY 19 -----FOAHO---- 23
Db 4738 AKOPAAARIAHISLAFDASSWEIYAPLLNGGAIVCADIYFTTIDPQALQETFOEHEIRGA 4797
QY 24 ----- 23
Db 4798 MLPPSLKQCLVOAPDWIGRLDILFAAGDRFSSVDALQORLVGSGVFNAYGPTENTILS 4857
QY 24 ----- 23
Db 4858 TIYNVAENDSFVNGVPIGSVNSGAYIMDKNOQLVPAGVMGELVVTGDLARGYMDPKL 4917
QY 24 ----- 23
Db 4918 DADRFIQLTVNGSEQVRAVYRTGDRVYRPKDFQIEFFGRMDQOIKIRHRIEPAEVEQAF 4977
QY 24 ----- 23
Db 4978 LNDGFVEDVAIVRTPENQEPWVAFVTAKGDNSAREEATTOIEGWEAHFEGGAYANIE 5037
QY 24 -----EDT----- 26
Db 5038 EIESEALGYDFMGWTSNYDGTEDKDEMREWLNDTMRSLLDGKPAGRVLEVTGTGTGMIMF 5097
QY 27 -----ERYV----- 30
Db 5098 NLGRSQGLERYIGLEPAPSAEAFVNNAAKSFPGLAGRAEVHVGTAAADVGTLOGLTSDMAV 5157
QY 31 ----- 30
Db 5158 INSVAQYFPTPEYLAETIKSLVQVPMKRIYVLGDMRSMWNRDFAAARAAYSADNASKD 5217
QY 31 ----- 30

Db 5218 RVROKMELEBKEBELLVDPAFFTALASQLQDRIQHVEILPKRMKATNELSSRYAAVLH 5277
QY 31 ----- 30
Db 5278 ISDEPLPIYKIDPEAWINFEGSRLTRELAQVLKENENAEVAISNIPYSKTVVERHIVR 5337
QY 31 ----- 30
Db 5338 SLDOEDANAPESMDGSDWISAVRTRAQOCHTSLASDLFIDAEDAGFRVSVWARQHSOH 5397
QY 31 -----LTN----- 33
Db 5398 GALDAVFHHLXPATEDSRVLKIFPTDHQGRPLKSLTNOPLLPAOSRRRAELLIREGLQTL 5457
QY 34 ----- 33
Db 5458 PPYMIPOITLIDRMLNANGKVDRELARAKITQSKPVEDIVPPRNSVEATVCKGFT 5517
QY 34 ----- 33
Db 5518 DVLGVEVGITDNFFNLGHSMLATKLAARLGRQLNTRISVRDVFDPQVYVADLAIVORNS 5577
QY 34 -----LNTGA----- 38
Db 5578 APHEPIKPADYTGVPVQSFQAQRLWFLDQLVNGATWYMLPLGIRLHLSRVDALATAISA 5637
QY 39 ----- 38
Db 5638 LEORHEPLRTTFHEEDGVGVQVQDHRPKDLRIIDLSTQPKDAYLAVLKHEQTTFLDLAT 5697
QY 39 ----- 38
Db 5698 EPGMRVALIRLGEHEHILSIVMHHIISDQWSVEVLDFEMHRFPYSALRQODPMEQILPLP 5757
QY 39 ----- 38
Db 5758 IOYRFAAAQKTEQVAEHOQOLDYVTEHLADSTPAELLTDLPRPSILSGRANELPLTIE 5817
QY 39 ----- 38
Db 5818 GELHDKLRAFCRVHQAOTPFVILLAAALRAAHYRLTGAEDATLGTPIANRNRPELENNIGFF 5877
QY 39 ----- 38
Db 5878 VNTQCMRIAENDNFESLVRVRSTATSAFANQDVPFESIVSSLLPGSRDASRNPLVQV 5937
QY 39 ----- 38
Db 5938 ILAVHSQDQDLGKLTLEGLRDEAVDSAISTRFDVEPHLFEHADRLSGSVLYAKELFKLRTI 5997
QY 39 ----- 38
Db 5998 ESVVSVFLETLRRALDQPLTPLAVLPLTDGVCGEIASKGLDVPRTDYPRDANIVEVFOOH 6057
QY 39 ----- 38
Db 6058 VRATPDIAVKDATSILTVAQLDQSDRLAIWLSRRHMPETLVGLVAPRSCETIIAMFG 6117
QY 39 ----- 38
Db 6118 IMKANLAYLPLDINSAPAARLRSILSAVDGNKVLVLLGSGVTAPEQENPEVEAVGICQILAG 6177
QY 39 ----- 38
Db 6178 TGLDKTQGSNARPSATSLAYVIFTSGSTGKPGVMVHRSVTRLAKPSNVISKLPOGARV 6237
QY 39 ----- 38
Db 6238 AHLANIAFDASIEIATILLNGATLVCLDYHTVLCRTLKEVFRESITVTLMPALLKQ 6297
QY 39 ----- 38

```
Db 6298 CVABIPETLAHLDLLTYTGGDRVGGHDAMRARSVLKIQMFGSYGPTENTVISTIEYVDADE 6357
Qy 39 ----- 38
Db 6358 MFVNGVPIGKTVNSGAYVMDRNOQLVPSGVVGLVVTGDLARGYTDPSLKNKRFIYT 6417
Qy 39 ----- 38
Db 6418 VNGESIRAYRTGDRVYRPHDLQIEFFGRMDQVKIRGHRIEPGEVESALLSHNSVQDAA 6477
Qy 39 ----- 38
Db 6478 VVICAPADQDSGAEMVAFVAARNTEDTQEEBAVDQVQWETHFETAAYSEVKDIRQSE 6537
Qy 39 ----- 38
Db 6538 VGNDFMGWTSWYDCSEIDKTDHWEHLNDTMRMILDAREPGHVLGIGTGMVMFNLAACP 6597
Qy 39 ----- 38
Db 6598 GLQYVGPEPSKAAQFYNDAAQSPALKDGRSIVHVGATDINKAGPIQPLVVINSVA 6657
Qy 39 ----- 38
Db 6658 QYFPTPEYLFVRVEALVOIPSEVERIVFGDMRTNAINRDFVASRALHTLGEKANKELVRQM 6717
Qy 39 ----- 45
Db 6718 IYELEANEELLTOPAFTSLRTRLGEKIKHVEILPKTMKATNELSKVRYAAVLHVRGSR 6777
Qy 46 ----- 45
Db 6778 EQSTHIVSPNAWIDFAADGLDROTLINLLKEHKDAGTVAIGNIPYSKTIVERFVNKSLS 6837
Qy 46 ----- 45
Db 6838 EDDMEEGQNSLDGSAWVAARMAAQCPSLDAMDYKEITAQAGYQVEYSWARQWSONCAL 6897
Qy 46 ----- 49
Db 6898 DAIPIHFEPPEKEGARTLIEFTDYEGRNVTNLTNRPLNSIOSRRIGTQIREKLOTLPPY 6957
Qy 50 ----- 49
Db 6958 MIPSRIMVLDQMPVNNNGKIDRKELVRRRAIVAPKPSAATRVAPRNEIEAILRDEFEDVL 7017
Qy 50 ----- 49
Db 7018 GTEVSVLDNFFDLGGHSLMATKLAARVSRRLDAHISIKDVDPQVPLADLAASIORESAPH 7077
Qy 50 ----- 49
Db 7078 EPIQRPYTPAEQSAQGLWFLDQNLGATWYLMPLAIRGQLRVAALSAAALFALER 7137
Qy 50 ----- 49
Db 7138 RHETLRTTFEESGQVQIVGEARNSDLRVHDVSTGDDGEYLEVLRREQTVFDFLSSEPG 7197
Qy 50 FRVHLVK----- 56
Db 7198 WRVCLVKTGEEDHVLIVMHHIYDGMVSDILRGELQGFYSAAALRGQDPLLHANPLPIQY 7257
Qy 57 ----- 56
Db 7258 RDPAAWREAKQVEEHQRQLGYWSKQVDSPTPAELLTDLPSPILSGRAGSDVDTIEGSV 7317
Qy 57 ----- 61
Db 7318 YGALQSFCTRSTVTFVLLTVFRIAHFRLTAVDATIGTPIANRNPPELETILVGCFVNT 7377
Qy 62 ----- 61
Db 7378 QCMKRISTADDNFEGLVRQVRNVATAAYANQDVPFERIVSALVPGSRNTSRNPLVQLMFA 7437
Qy 62 ----- 61
Db 7438 VOSVEDYDQVRLEGLSVNMPGEASTRFDMEPHLVPGDOKLTGSLVYSSDLFEQGTIONF 7497
Qy 62 ----- 61
Db 7498 VDIFQECLSVLDQPLTPISVLPSNAISNLESOLDLEMPSTDYPRDRTVVDLFREQAAI 7557
Qy 62 ----- 61
Db 7558 CPDSIAVKOSSQLTYAQLDEQSDRVAWLHERHMPAESLVGLVSPRSCETIIAYFGIMK 7617
Qy 62 ----- 61
Db 7618 ANLAYPLDYAPDARLAAILDITVEGERLLLLGAGVQPQGIQIPRLSTAYIAEALSHATT 7677
Qy 62 ----- 65
Db 7678 VDVTSIPQPSATSLAYVIFTSGTGPKGMIEHGRGIVRLVRDTNVTNNVPESGALPVSH 7737
Qy 66 ----- 65
Db 7738 FSNLAWDAATWEIYTAVLNGTVCIDRDTMLDIAALNSTFRKENVRAAFTTAPFLKQCL 7797
Qy 66 --APNITANL----- 73
Db 7798 AETPELVANLEILHTAGDRDLDPGDANLAGKTAKGGIFNVLGHNTENTAYSTFYVPVGEETF 7857
Qy 74 ----- 73
Db 7858 VNGVPVGRGISHAYIIDRHOKLVPAGVMGELILTGDGVARGYDSALNKRPFYIIDIN 7917
Qy 74 ----- 73
Db 7918 GKSTWSYRTGDKARYRPRDQGLEFFGRMDQWVKIRGRIEPEGEVELTLLDHKSVLAATVV 7977
Qy 74 ----- 73
Db 7978 VRRPPNGDPEMIAFITIDAEDDVQTHKAIYKHLQGLPAYMIPSHLVILDQMPVTDNGKV 8037
Qy 74 ----- 73
Db 8038 DRKDLALRAQTOKRRSTAARVPRDEVAVLCEEYSNLLLEVEGITDGFDFLGGHSLLA 8097
Qy 74 ----- 73
Db 8098 TKLAARLSRQLNTRVSVKDVDPQILADLADIIRRGSHRHDPIPATPYTGPVEQSAQGR 8157
Qy 74 ----- 73
Db 8158 LWFLEQLNLGASWYLMPPFAIRMRGPLOTKALAVNALVHRHEALRTTFEDHDGVGVQVI 8217
Qy 74 ----- 73
Db 8218 QPKSSQDLRIIDLSDAVDDTAYLAALKREOTTAPDLTSEPGWRVSLRLGDDDDYILSIVM 8277
Qy 74 ----- 73
Db 8278 HHIISDGHVTVLRLQELQGFYSAAIRGQEPQSAKSLPIQYRDFAVMQRQENQIKQAKQ 8337
Qy 74 ----- 73
Db 8338 LKWSQQLADSTPCFELTDLPRPSILSCEADAVPMVIDGTVYQLLTDPCRTHQVTSFVSL 8397
Qy 74 ----- 73
Db 8398 LAAPFTHYRLTGLDATVGTPIANRNPPELEGLIGFFPWTQCMRMAISETETTESLVOQ 8457
Qy 74 ----- 73
Db 8458 VRLTTTEAFANQDVPFEQIVSTLLPGSRDTSRNPVQVNFALQSQOQLGRIOLEGMTDEA 8517
```

QY 74 ----- 73
Db 8518 LETPLSTRLDLEVLHPOEVCKLSGSLLYSTDLEFEVETIRGIIVDVDFLEILRRGLEQPKQRL 8577
QY 74 ----- 73
Db 8578 MAMPITDITKLRDOGLLTAKPAYPRESSVIDLFRQVAAAPDAIAVWDSSTLTLYADL 8637
QY 74 ----- 73
Db 8638 DQOSNLAHLWCORNAPETLVAFAPRSCLTIVAFGLVKANLAYPLDVNAPARIEA 8697
QY 74 ----- 73
Db 8698 ILSAVPGHKLVQAHGPELGLTMADELVOIDEALASSSGDHEQIHASGPTATSLAVV 8757
QY 74 ----- 73
Db 8758 MFTSGTGPKGVMIDHRSIIRLVKNSDVVATLPTVRMANVSNLAFDISVQBIYTALLN 8817
QY 74 ----- 73
Db 8818 GGTLVCLDYLLDLSKILYNVFEAQVNAAMFTPVLLKQCLGNMPAIIIRLSVLFVNGDR 8877
QY 74 ----- 73
Db 8878 LDAHDAVASGLIODAVYNAYGPTENGCMQSTMYKVDVNEPFVNGVPIGRSITNSGAYVMD 8937
QY 74 ----- 73
Db 8938 GNOQLVSPGVMEIVVTGDGLARGYTDALDEDRFVHVTTIDGENIKAVRTGDRVYRPK 8997
QY 74 ----- 73
Db 8998 DFEIEFGRMDQVKIRGRIEPAEVEHALLGHDLVHDAAVLVRKANPEMIAFITSQ 9057
QY 74 ----- 73
Db 9058 EDETIEQESNKVQVQGEHFDVSRYADIKDLDTSTFGHDFLGWTSMDYDGDIPVNEKME 9117
QY 74 ----- 78
Db 9118 WLDETTASLLDNRPBGHILEIGAGTGMILNGLKVDGLQYVGLDPAPSAAIFVNEAVKS 9177
QY 79 ----- 78
Db 9178 LPSLAGKARVLVGTALDIGSLDKNEIOPELVVINSVAQYFPTSEYLIKVKVAVVEVPSVK 9237
QY 79 ----- 78
Db 9238 RVFGDIRSQALNRDFLAARAVRALGDNASKEQIREKIAELESEEEELLVDPAFFVSLRS 9297
QY 79 ----- 78
Db 9298 QLPNIKHEVLPKLMKATNELSYAAVLHISHNEEQLLIODIPTAWDFAATOKDS 9357
QY 79 ----- SVC 81
Db 9358 QGLRNLLQGRDDVMIAVGNIPYSKTIVERHIMNSLDQDHVNSLDGTSWISDARSAAC 9417
QY 82 ----- 81
Db 9418 TSPDAPALTOLAKEGFRVELSWARQSONGALDAVFHRLATDANCERSRVLVHFPTDHO 9477
QY 82 ----- 81
Db 9478 GRQLRTLNRPLQASRRIBESQVFEALQALPAYMIPRSRIIVLPQMPTNANGKVDKQOL 9537
QY 82 ----- 81
Db 9538 ARRAQVAKRKAVSARVAPRNDTEIVLCEEYADILGTEVGITDNFFDMGGHSLMATKLA 9597
QY 82 ----- 81

Db 9598 RLSRRLDTRVTVEVFDKPVLAADLAASIEQGSTPHLPIASSVYSGPVEQSYAQRLWFELD 9657
QY 82 ----- 81
Db 9658 QFNLNATYHMSLAMRLLGPLNMDALVALRALBORHETLTTFEAKDKIGVQVVHEAGM 9717
QY 82 ----- 81
Db 9718 KRLKVLDSLDKNEKHEMAVLENEQMRPFTLASPEQWKHGLARLGPTEYILSLVMHHMPSD 9777
QY 82 GWS ----- 84
Db 9778 GMSVDILROELGQFVSAALRGDRDPLSQVKPLPIQYVRDPAAMQKAAQVAEHERQLAYWEN 9837
QY 85 ----- 84
Db 9838 QLADSTPGELLTDFFRPQFLSGKAGVIPVTIEGPVYEKLLKFSKERQVTLFSLVLTAFRA 9897
QY 85 ----- 84
Db 9898 THFRLTGAEDATIGTPIANRNPELEHIIGFFVNTQCMRLDLDGTSTFESLVQHVRSVAT 9957
QY 85 ----- 84
Db 9958 DAYSQDIPFERIVSALLPGSRDASRPLIQLMFALHSQDPLGNITLEGLEHERLPTSVA 10017
QY 85 ----- QTIN ----- 88
Db 10018 TRFDMFHLFQEPNKLSGSILPADELFOPETINSVVTVQBEILRRGLDQPOVSISTMPLT 10077
QY 89 ----- 88
Db 10078 DGLIDLEKLGLEIESSNPPRDYSVVDVFRQQAANPNAPAVVDSETSMYSYTSLOKSEO 10137
QY 89 ----- PE ----- 90
Db 10138 IAAMLHAQRLRPESLICVNAPRSFETIVSLFGILKAGYAYLPDVNSPAARIQIPLSEVE 10197
QY 91 ----- 90
Db 10198 GKRLVLLSGIDPMOSDRMDVETARIQDILTNTKVERSPPMSRPSATSLAYVFTSGSTG 10257
QY 91 ----- 90
Db 10258 RPKGVMIHRNLRVLYKQSNVTSQLPQDLRMAHISNLPADASIWEIFTAILNGGALICID 10317
QY 91 ----- 90
Db 10318 YFTLLDSQALRTTFEKARVNTLFPALLKECLNHAPTLFEDLKVLYIGGDBLDATDAAK 10377
QY 91 ----- 90
Db 10378 IQALVKGTVYNAYGPTENTVMSTIYRLTDGESYANGVPIGNVAVSSGAYIMDOKORLVPP 10437
QY 91 ----- 90
Db 10438 GVMGELVWSDGLARGYTNSTLNADRFVDIVINDOKARAYRTGDRTRYRPKDGSIEFFGR 10497
QY 91 ----- 90
Db 10498 MDQOVKIRHRVEPAEVEQAMLGNAIKAHDAAVVQAVDQGETEMIGFVSNASDRPSEGE 10557
QY 91 ----- DOT ----- 93
Db 10558 EITNQVQEMEDHFESTAYAGIEAIDQATLGRDFTSMTSMYNGNLIDKAEMEELDDTMS 10617
QY 94 ----- 93
Db 10618 LLDKEDARPCAEIGTGTGMVLFNLNPKDGLSVYIEPSSAALFVDKAAQDPFGLOGKT 10677
QY 94 ----- 93

Db 10678 OILVGTAEIKLVKDFHPD VVINSVAQYPPSRSYLVQIASELIHMTSVKTIFFGDMRSW 10737
Qy 94 -----DP----- 95
Db 10738 ATNRDFLVSRALYTLGDKATKDOI RQEVARLEENEDELLVDPAPFTSLTSQWPGVKVHE 10797
Qy 96 ----- 95
Db 10798 ILPKMRMSTNELSSRYAAVLHICRDGEGRRYGRVHSVEENAWIDFASGMDRHALVQ 10857
Qy 96 ----- 95
Db 10858 MLDERRDAKTVAGNIPHSNTINERHFTTSLDTEGEGIAQDSLDSGSAWQSAWKAAWARP 10917
Qy 96 ----- 95
Db 10918 CLSVTELVEIGQAAGFRVSWARQSRQH GALDVVFHLEDDRVGRVLINPPTDFERLPP 10977
Qy 96 ----- 95
Db 10978 STGLTSRPLORIONRRFESQIREOLTLPPYMPVPSRIVVLERMPLNANSKVDKRELARK 11037
Qy 96 -----GH----- 97
Db 11038 ARTLOTIKPSATRVAPRNDIEAVLCDEFOAVLGVTVGMDNFFELGGHSLMATKLAARLS 11097
Qy 98 -----ADLV----- 101
Db 11098 RRLDTRSVSKDI FNOPILODLADVVTGSAPEAI PSTPYSGPVEQSFSGRLWFLDQLN 11157
Qy 102 ----- 101
Db 11158 LNASWYHMLASRLRGPLRIEALOSALATIEARHESLTTTFFEOGVPVQIVRAARNKQL 11217
Qy 102 ----- 101
Db 11218 RIIDVSGTEDAYLAALKOEQDAADF LTAEPGWRVALLRLGPDHVLVSIWMHHIISDGWSV 11277
Qy 102 -----LY----- 103
Db 11278 DILROELGOLYSNASSQAPLPIQYRDFAIWOKDSQIAEHQKOLNYWKQLVNSKPAEL 11337
Qy 104 ----- 103
Db 11338 LADFTPRKALSGADVPIEIDDQVYQNLRSFCRARHVTFSVALLAAPRAAHYRLTGAED 11397
Qy 104 ----- 103
Db 11398 ATIGSPIANRNPPELEGLIGCFVNTQCLRI PVKSEDTFTLLVKQARETATEAQDNQDVPF 11457
Qy 104 -----ITRFDLEL----- 111
Db 11458 ERIVSSMVASSRDTSRNPLVQWFAVSHQDLGNIRLEGVEGKPVSMASSTRFDEAMHLF 11517
Qy 112 ----- 111
Db 11518 EDQGMGGNVFVKDLFESETIRSVAVPQETLRRGLANPHANLATLPLTDCGLPSRLSLC 11577
Qy 112 -----PD----- 113
Db 11578 LQVNPQDYPDRDASVIDVFEQVASIPKSTAVIDASSQLTYTELDERSSQLATWLRQVTV 11637
Qy 114 -----GNRQV----- 118
Db 11638 PEELGVGLAPRSCETIIAFLGIKANLAYLPLDVNAPAGRIETIILSSLPFNGRLILLGSDT 11697
Qy 119 -----R----- 119
Db 11698 QAVKULHANSVRFRTRISDALVESGSPTELSRTPAQSLAYVMFTSGTGVKPGVMVHR 11757
Qy 120 GVT----- 122
Db 11758 GITRLVNSNVAKOPAAAAIAHLSNIAFDASSWEIYAPLLNGGTWVCIDYVTTIDIKAL 11817

Qy 123 ----- 122
Db 11818 EAVFKOHHRGAMLPALLKQCLVSAPTWMISSEILFAAGDRLLSSODAILARRAVGSGVY 11877
Qy 123 ----- 122
Db 11878 NAYGPTENTVLSTIHNIGENEA FNGVPIGNAVSNGSAFMDQOQLVAGVIGELVVVTG 11937
Qy 123 ----- 122
Db 11938 DGLARGYTD SKLRVDRFIYITLDGNNRAYRTGDRVHRPKDGGQIEFFGRMDOQIIRGH 11997
Qy 123 ----- 122
Db 11998 RIEPAVEQALARDPAISDAVITQLTDEEPELVAFPSLKGNGANGTVNGVSDQEKID 12057
Qy 123 ----- 122
Db 12058 GDEQHALLMENKIRHNLQALLPTYMIPSR IIHVDQLPVNANGKIDRNELAVRAQATPRTS 12117
Qy 123 ----- 122
Db 12118 SVSTYVAPRNDIETIICKEFADILSVRVGITDNFFDLGGHSLIATKLAARLSRRLDTRVS 12177
Qy 123 ----- 122
Db 12178 VRDVFDPVVGQLAASIQQGSTPHEAI PALSHSGPVQQSPAQGRLWFLDRFNLAAMYIM 12237
Qy 123 ----- 122
Db 12238 PFGVRLRGPLRV DALQTALRALEERHELLRTTPEEQDGVGMQIVHSPMRDICI VWDISGA 12297
Qy 123 ----- 122
Db 12298 NEDLAKLEBQQAPFNLSTEVAMRVALFKAGENHHILSI VMHHIISDGWSVDI PQBELAQ 12357
Qy 123 ----- 122
Db 12358 FYSVAVRGHDP LSOVKPLPIHYRDFAVWQRQDKQAVHESQLQWIEQLADSTPAELISD 12417
Qy 123 ----- 122
Db 12418 FNRPEVLGSEAGTVPIVIEDEVYEKLSLFCRNHQVTSFV VLLAAFRVAHYRLTGAEDATI 12477
Qy 123 ----- 122
Db 12478 GTPIANRNRPELEDLIGFFVNTOCMR IALEEHDNFLSVVRVRSTAASAFENQDVPFERL 12537
Qy 123 ----- 122
Db 12538 VSALLPGSRDASRNPLVQLMFVVHVSORNLCKLQLEGLEGEPTPYTATTRFDVEPHLFEQD 12597
Qy 123 ----- 122
Db 12598 KGLAGNVVFAADLFEAATIRSVVEFPHEILRRGLDQDPDIAISTMPLVDGLAALNSRLNLA 12657
Qy 123 ----- 122
Db 12658 VEDIEPDPATEASVVVVFQTVQVNPALAVTD TSTKLTVAELDQOSDHVAWLKOKLP 12717
Qy 123 ----- 122
Db 12718 AESIVVVVLA PRSSETIVACIGILKANLAYLPMDSNVPEARQAILSEIPGEKPVLLGAGV 12777
Qy 123 ----- 122
Db 12778 PIPDNKTADV RMVFI SDIVASKTDSYSGTPRSASSLAYVIFTSGTGRPKGMVMEHRG 12837
Qy 123 ----- 122
Db 12838 VISLVKQNASRIPOSLRMAHVSNLAFDASVWEI FTILLNGGTLFCISYFTVLDSKALSAA 12897

QY 123 ----- 122
Db 12898 FSDHRINITLLPALLKQCLADAPSVLSLESYIGGDRLDGADATKVKDLVKGKAYNAY 12957
QY 123 ----- 122
Db 12958 GPTENSVMSTIYTIHETFANGVPIGTSGLGPKSKAVIMDDQQLVPAGVMGELVWAGDGL 13017
QY 123 ----- 122
Db 13018 ARGYDPSLNTGRFIHITIDGKQVAYRTGDRVRYRPRDYQIEFFGRLDQOIKIRHRIE 13077
QY 123 ----- 122
Db 13078 PAVEQALLSDSSINDAVVVSQAQKGLEVMGYITTOAAQSVDKKEASNKVQWEAHFDS 13137
QY 123 ----- 122
Db 13138 TAYANIGGIDRDALGDQFLSWTSYDGLIPREEMQEWLNDTMRSLDNPQPGKVLIGT 13197
QY 123 ----- 122
Db 13198 GTGMVFLNGLKVEGLOSAGLEPSRSVTAWNKAIEFPFSLAGSARVHVGTAEIDISSDG 13257
QY 123 ----- 122
Db 13258 LRSDLVINSVAQYFFPSREYLAEITANLIRLPGVKRIFFGDMRTYATNKDFLVARAVHTL 13317
QY 123 ----- 122
Db 13318 GSNASKAMVQVQAKLEDEELLVDPAFTSLSDQFPDEIKHVEILPKRMAATNELSSY 13377
QY 123 ----- 122
Db 13378 RYAAVHVGHGHPNGEDEDKQWAVKINPKAWVDFAGTRMDRQALLQLLQDRGDDVV 13437
QY 123 ----- 122
Db 13438 AVSNIPYSKTIMEHLSQSLDDDEGTSVDGTAWISRTOSRAKECPALSVADLIEIGKG 13497
QY 123 ----- 122
Db 13498 IGFEVEASWARQHSQRGLDAVFRHPPRHSGHVMFRFTEHKGRSSSLTNRLHLQ 13557
QY 123 ----- 122
Db 13558 SRLEAKVRERLOSLLPPYMIPSRITLLDQPLTNSGVKDRKKLARQARVIPRASAATLD 13617
QY 123 ----- 126
Db 13618 FVAPRTEIEVVLCEFTLLGVKVGITDNFFELGGHSLLATKLSARLSRLDAGITVKQV 13677
QY 127 ----- 126
Db 13678 FDQVPLADLAASILQSSRRHSRISPLPYEGVEQSAQGLWFLDQFNIDALWYLIIPAL 13737
QY 127 ----- 126
Db 13738 RMRGFLQDALAAALVALEERHESLRTTFRDGVGVQVQPLRTTKDIRIIVSGMRDD 13797
QY 127 ----- 132
Db 13798 DAYLEPQKEQQTFFDLASEPGWRVALLKLGKDDHILSVHWHIISDQWSTEVLORELQ 13857
QY 133 ----- 132
Db 13858 FYLAASKGAPLSQVAPLPIQYRDFAVWQROEQVAESQRLDVKWKLADSPAEILLAD 13917
QY 133 ----- 132
Db 13918 YTRPNVLSGEAGSVFVINDSVYKSLVFCRSRQVTTFTTLAAFRAAHYRMTGSDDATI 13977
QY 133 ----- 136

Db 13978 GTPIANRNPELENLIGCFVNTQCMRITIGDDETFESLVQOVRSTTATAFENQDVPFERI 14037
QY 137 ----- 144
Db 14038 VSTLSAGSRDTSRNLPLVQLLFAVHSQOGLGRIQLDGVVDPEVLSTVSTRFDLEFHAFOEA 14097
QY 145 ----- 144
Db 14098 DRLGNSVMFATDLFQPETIQGFVAVVEVLQRLGEPQSPQSIATMPLAEGIAQLRDAGALQ 14157
QY 145 ----- 144
Db 14158 MPKSDYPRNASLVDVFQOAMASPSTVAVTDSKLTLYAELDRLSQQAASYLRRQOLPAE 14217
QY 145 ----- 144
Db 14218 TWAVLAPRSCETIIAFLAILKANLAYMPLDVNTPSARMEAISSVPGRRLLVLVSGVRH 14277
QY 145 ----- 144
Db 14278 ADINVPNAKTMLISDVTGTDAIGTPEPLVVRPSATSLAYVIFTSGSTGKPKGMVMEHRA 14337
QY 145 ----- 144
Db 14338 IMRLVKDSNVVTHMPPTATRMHAHVNTIAFDVSLFEMCATLLNGTGLVCIDYLLDSTMLR 14397
QY 145 ----- 144
Db 14398 ETFEREQVRAAIFPPALLRQCLVNMPPDAIGMLEAVYVAGDRFHSRDARATQALAGPRVYN 14457
QY 145 ----- 144
Db 14458 AYGPTEAAILSTIYNIDKHDPYVNGVPIGSAVNSGAYVMDRNQQLPPGVGELVVTGE 14517
QY 145 ----- 144
Db 14518 GVARGYTDASLTDREVTVTIDGQORAYRTGDRVYRPRKGFQIEFFGLDQOAKIRGHR 14577
QY 145 ----- 144
Db 14578 VELGEVEHALLSENSVTDAAVVLRTWEEDPOLVAFVTTDHEYRSGSSNEEDPYATQAA 14637
QY 145 ----- 144
Db 14638 GDMRKELRSLLPYVMPSRVITILRQMPLANGVKDRKDLARRAQMTPTASSSGPVHVAPR 14697
QY 145 ----- 147
Db 14698 NETEAAICDEFETILGVKVGIT 14719

RESULT 3

T30192

probable peptide synthetase - Aureobasidium pullulans

C:Species: Aureobasidium pullulans

C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000

C:Accession: T30192

R:Peery, R.B.; Thornnewell, S.J.; Tobin, M.B.; Skatrud, P.L.

submitted to the EMBL Data Library, January 1997

A:Description: Discovery of an MDR-like gene adjacent to a peptide synthetase in Aure

A:Reference number: Z20767

A:Accession: T30192

A>Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-10797 <PEE>

A:Cross-references: EMBL:U85909; NID:G4099310; PID:G4099313; PIDN:AAD00581.1

C:Genetics:

A:introns: 2078/3; 4142/3; 6075/3; 7963/3; 7985/2

C:Superfamily: acyl carrier protein homology

C:Keywords: carrier protein

F:1618-1688/Domain: acyl carrier protein homology <ACP1>

F:3682-3752/Domain: acyl carrier protein homology <ACP2>

F:5615-5685/Domain: acyl carrier protein homology <ACP3>
F:7503-7573/Domain: acyl carrier protein homology <ACP4>
F:9683-9752/Domain: acyl carrier protein homology <ACP5>

Query Match 31.6%; Score 244; DB 2; Length 10797;
Best Local Similarity 1.1%; Pred. No. 55;
Matches 94; Conservative 27; Mismatches 27; Indels 8749; Gaps 30;

```
QY 1 AAG----- 3
   |||
Db 575 AAGDSLRAIQVASQLRQGCADLCALLESVSNVLALDELVRKADESNINDNDSEKWK 634
QY 4 ----- 3
Db 635 QTVHELRSVERDPESKAFVSGIEDVLPCTPLQDAMLVETAKRPOAYCNELRLLTVSPKIP 694
QY 4 ----- 7
   GILH-
Db 695 VERVRQALFALAQRHTALRSQFMPSGVSHCAVTVQVIMKTLVTSQFAHVKSFTTCGWSVNR 754
   |||
QY 8 ----- 13
   -LELIVA-
   |||
Db 755 ETLLRPLHFQYKCGAEAEILVAIHIALYDQMSVEVILELTLQNERTPERPSFGAVN 814
QY 14 ----- 13
Db 815 KFFNLRRSEDQSHLDFWGEYLSVDVTPGRPLNLSPPMMPPPOPLQSIQHTIEMDMETLROA 874
QY 14 ----- 16
   -VGP-
   |||
Db 875 AHSYSCSAHVFOAYAILGFYMGTEDTVFGTVFSGRTLPIVEIESVGVPLLSLTPRI 934
QY 17 ----- 25
   DVFOAHQED-
   |||
Db 935 NTLESRKFSVLSRLQEDNRKIMRHSMTSLADIKKACGFNGEAVFDSIFVWQETARPD 994
QY 26 ----- 25
Db 995 RAQTLNLVLAHDYLFNLTLEPTQQGVKTKATYQSSLLPLQHVKTLLQOLDALXIV 1054
QY 26 ----- 25
Db 1055 VARPETHNNEISDQLPISVLSVANSEPOSFVYKAGLSGVNHALNNSGGLALVFAHIR 1114
QY 26 ----- 25
Db 1115 ECTSRMESLTYGELNTRANQLANYLISQAKRDELICVMEKSVSLYLILAAVKAGCGY 1174
QY 26 ----- 25
Db 1175 LPLVPETPAARIRQILAEADVVKFCLTDSMAPVIAADVSRCHIMNVDTTDCSAQSGTGPQL 1234
QY 26 ----- 36
   -TERYVLTNLNI-
   |||
Db 1235 DFKPTDIAVAVTSGTTGKPKGLVLTQENILSNLEVLKSIYVPVPEGSRLQACNQAFDVS 1294
QY 37 ----- 36
Db 1295 VFEIIFTWTYGMCLCSASKVDVFRDFEKAINELEITHLSLTPTVAALTPAHIPRVKFLV 1354
QY 37 ----- 36
Db 1355 TAGEAVTHHVHGAWAKGLYQGYGPSETTNICTVNSAVESDHVINNIGPAFENTSFAVLT 1414
QY 37 ----- 36
Db 1415 QGDDFQLVPLGGLGELFCGQQVFRGYQNNPELTESKIINHPNYGRIYRSGDLGRLLPDG 1474
QY 37 ----- 36
Db 1475 TILIOGRDDBKIRGORIELGEISGCLLQFPSPVONCAIEVIKTADEKRLMAFWIPSGYS 1534
QY 37 ----- 36
```

```
Db 1535 KDSYSILOPKNLEEIIKSIYAHADNLPAYMVPDALVPVSAIPOTSQKIDKRRRLASDG 1594
QY 37 ----- 36
Db 1595 SALTVEDLNAYSRGADDDTSELSATEOOLASALADTLOMSQTSIGRSTSFALGLDSVS 1654
QY 37 ----- 36
Db 1655 AIRLATNLRKEYGYSIDVSQILKRPTIARLAPLLGGESSKQTNEPVTADCEAAVGSYLHD 1714
QY 37 ----- 36
Db 1715 ESVVSQLHEHGQTVSQVLPTQLEAMLSARDTSGSSAYRNKTLFSLHGSVDKLKACWEV 1774
QY 37 ----- 36
Db 1775 MLORHDILRTIFLSTEDSRFPFVOAVLSQWTLPWQECDDIPDQLSTLLDSAKAGGDSIVD 1834
QY 37 ----- 36
Db 1835 HSPPWKIQVYRSESTVYLLDMHMHALYDANAMSNLLYEQLYKDQSLSAVPSFKPFLNF 1894
QY 37 ----- 36
Db 1895 MISTSVBEADALPDQLREFVPKPKRTOVKSGFGTITGRNLNYSPKMVETFLSKHSTML 1954
QY 37 ----- 36
Db 1955 SLTQAMMMKTLAASQSYSDVCCGVNVSGRSPVVDGIESLVAPCPNTIPVRVDLSKHSNL 2014
QY 37 ----- 36
Db 2015 GLVKALQRVNIDSLPYQLTPLRRIQAAGTNGKRLPDSLVLLOQDITDLSAIWRLEGES 2074
QY 37 ----- 43
   -GAELLRD-
   |||
Db 2075 GVMDDETLSILNPLPLLEGPELLHELISGPETSGTALEYLDADGKTITLSYLDLHRRS 2134
QY 44 ----- 43
Db 2135 EHLAWRLIEVSRHISRKDLGIVPIYIPQCISLYISQALIKSGAAFCPLNLDVPGERLKF 2194
QY 44 ----- 43
Db 2195 ILKDTASILLTTAAMRPKLPELEGITVIIVDDDELSEFKSDREPAGEKITSPDTSLAYI 2254
QY 44 ----- 43
Db 2255 MYTSGSTGLPKAVCLSHRAVTQSLLAHDFIPSPSRFLOFASPTFDVSFPPWYRGA 2314
QY 44 ----- 43
Db 2315 TLVSVERNRLGDLPGTITSLNIDAAELTPSVAASLVRRHRENVPTLRALLTIGEMLNTQV 2374
QY 44 ----- 43
Db 2375 IQFGGSGIKSGILYMGYPTEAAIHCTLOPGFVDLPAGTIGIPLDVTSCFIVKPTST 2434
QY 44 ----- 51
   -PSLGAQFR-
   |||
Db 2435 KHASOLEILPIGEIGELVIGGHQADGYNREEQTRAAFVTHPKFGGLYRTGDKARLHRN 2494
QY 52 ----- 51
Db 2495 GTLECYGRISSGQVKLRQORVELGEIEHAASKAGGCHAVIASVISGLLVLCIGDPHRVS 2554
QY 52 ----- 51
Db 2555 SKDIKSACQKWLPAYMIPSDIVLLDDFPYLPSPGVKDKKRLTDYNSNTAQHVSGSSDLSE 2614
QY 52 ----- 51
```

Db	2615	NAREITRIIESVLGVSIDHSTDLSSAAGLSLRAIQVASQLRRQCCADLGALELLSVSNVL	2674
Qy	52	-----VH-----	53
Db	2675	ALDELVRAKADESNINDNDSEKWKQTVHELRSRVERDPESKAFVSGIEDVLPCTPLQDAM	2734
Qy	54	-----	53
Db	2735	LVEAKRPOAYCNELRLTVSPKIPVERVRQALFALAQRHTALRSGFMPSGVSHCAVTQVI	2794
Qy	54	-----	53
Db	2795	WKTIVTSQFAHVKSFTTGWSVTNRETLRLPLHFQYKCSGAEAEILVAIHALYDQMSVEV	2854
Qy	54	-----	53
Db	2855	ILEDLETLNQNERTPERPSFGAVNKFNLNRSEDOTSHLDFWGEYLSDVTPGRLPNLSPK	2914
Qy	54	-----	53
Db	2915	MMPPOQLQSIQHTIEMDMETLROAAHSYSCSAHVFFQAAVAAILLGFYMGTEDTVFGTVFS	2974
Qy	54	-----	53
Db	2975	GRTLPIVEIESVMGPLLSTLPTRLTLESRRKFSVDLSRLQEDNKRIMRHSMTSLADIKKA	3034
Qy	54	-----	53
Db	3035	CGFNPGEAVFDSIFVMOETARPDAQOTLLNLVEAHDYLFENLTLELEPTQOGVKTATY	3094
Qy	54	-----LVXWVI-----	59
Db	3095	QSSLLPLQHVKTLLQQLDALVKIVVARPETHMNEISDQLPISVLSVANSEPSQSVYKAGL	3154
Qy	60	-----	59
Db	3155	GSLVENHALNSSLALVFAHDIREGTSRMESLTYGELNTRANQANLYLISQGAKRDELI	3214
Qy	60	-----LTS-----	62
Db	3215	CVCMEKSVLSLSILAAVAKCGYLPLVPETPAARIQILAEADVFCCLTDSSNAPVIAD	3274
Qy	63	-----	62
Db	3275	VSRCHINVDTTDCSAQSGTGPQLDFKPTDIAYAVFTSGTTGPKPKVLVTQENILSNLEV	3334
Qy	63	-----PEGA-----	66
Db	3335	LSKIYPVEGSRLLQACNQAFDVSVEIFFTWYTMCLCSASKDVMFRDFEKAINELEIT	3394
Qy	67	-----PNITANL-----	73
Db	3395	HLSLTPVVAALTDPAHIDPRVKFLVTAGEAVTHHVHGWAGKLYGGYGPSETTNICTVNS	3454
Qy	74	-----	73
Db	3455	AVESDHVINNIGPAPENTSFAVLITQDDDFQLVPLGLGELCFGGQVQPRGYQNNMPELTES	3514
Qy	74	-----	73
Db	3515	KIINHPNYGRIYRSGDLGRLLPDGTILIQGRTDDQKIRGORIELGEISGCLLOFPSPVQN	3574
Qy	74	-----	73
Db	3575	CAIEVIKTADKERLMAFWIPSGYSKDSYSILOPKNLEIISYIAHLADNLPAVMYVDA	3634
Qy	74	-----	73
Db	3635	LVPVSAIPQTSQKIDKRRLASDGSALTVEDLNAYRGADDDTSELSATEQOLASALAD	3694
Qy	74	-----	73
Db	3695	TLQMSQTSIGRSTSPFALGLDSVSAIRLATNLNRKEYGYSIDVSQILKRPTTARLAPLLGG	3754

Qy	74	-----	73
Db	3755	ESSKQTNBVTADCEAAVGSYLHDESVVSQLHEHQGTVSQVLPCTPQLOEAMLSARDTSGS	3814
Qy	74	-----	73
Db	3815	SAYRNKTLPSLHSGVDKXKACWEVMQORHDIILRTIFLSTEDSRFPFQVAVLSQWTLPHQOE	3874
Qy	74	-----	73
Db	3875	CDDIPDLSTLLDSAKAGGDSIVDHSPPMKIQVYRSESTVYLLLDHMHALYDANAMSNLL	3934
Qy	74	-----	73
Db	3935	YEVEQLYKDOSLAPVSPFKPFLNPMISTSVBEADALFRDQLREFVPKPKRTDVKSGFGT	3994
Qy	74	-----	73
Db	3995	ITGRNLNYSKQWVETFLSKHSTTMLSITQAMWMTLASOSYSYSDVCCGNVVSGRSVPVDGI	4054
Qy	74	-----	73
Db	4055	ESLVAFCFNTIPVRVDLSKHSRNLGLVKALQRWNIDSLPYQLTPLRRIQOAGTNGKRLF	4114
Qy	74	-----	73
Db	4115	DSLVLLOQDITDLSAIMRLEGESGYMDMRPKLPELEGITIVVDDDELSEFKSDREPAGE	4174
Qy	74	-----TSS-----	76
Db	4175	KITSPDTSLAYIMYTSGTGLPKAVCLSHRAVTQSLLAHDFRIFPSFSRFLQFASPTFDV	4234
Qy	77	-----	76
Db	4235	SVPEIFFPWRGATLVSVERNRLGLDLPCTITSLNIDAAELTPSVAASLVHRHENVPTLR	4294
Qy	77	-----	76
Db	4295	ALLTIGEMLNTVIOQFGSGICKSGILYMGYPTAAIHCTLOPGFVDLPAGTIGIPLD	4354
Qy	77	-----	76
Db	4355	TVSCFIVKPTSTYKSHASQLEILPIGEIGELVIGGHQADGLNLRBEOTRAAPFVTHPKFGG	4414
Qy	77	-----	76
Db	4415	LYRTGDKARLHRNGTLECYGRISSGQVKLRQORVELGEIEHAASKAGGCHAVIASVISGL	4474
Qy	77	-----	76
Db	4475	LVLFCIGDPHRVSSKDIKSACQKMLPAYMIPSDIVLLDDFPYLPSPGVKVKRLETDYNSN	4534
Qy	77	-----	76
Db	4535	TAOHVSGSSDLSENAREITRIIESVLGVSIDHSTDLSSAAGLSLRAIQVASQLRROSCAD	4594
Qy	77	-----LLSV-----	80
Db	4595	LGALELLSVSNVLDELVRAKADESNINDNDSEKWKQTVHELRSRVERDPESKAFVSGI	4654
Qy	81	-----	80
Db	4655	EDVLPCTPLODAMLVETAKRPOAYCNELRLTVSPKIPVERVRQALFALAQRHTALRSGFM	4714
Qy	81	-----	80
Db	4715	PSGVSHCAVTQVIWKTIVTSQFAHVKSFTTGWSVTNRETLRLPLHFQYKCSGAEAEILVA	4774
Qy	81	-----	80
Db	4775	IHHALYDQMSVEVILEDLETLNQNERTPERPSFGAVNKFNLNRSEDOTSHLDFWGEYLS	4834

Qy 81 ----- 80
Db 4835 DVTGRLPNLSPKMPPOPLQSIQHTIEMDMETLQAAHSYSCSAHVFFQAAVAILLGFY 4894
Qy 81 ----- 80
Db 4895 MGTEDTVEGTVFSGRTLPIVEISMVGPLLSTLPRINTLESRKFSVDLSRLQEDNRKIM 4954
Qy 81 -----CG-----WSOTINPE----- 90
Db 4955 RHSMTSLADIKKACGFNPGEAVFDSIFVQETARPDARAQTLLNLVEAHDYLEFNLTLEL 5014
Qy 91 ----- 90
Db 5015 EPTQGVKTKATYQSSLLPLQHWKTLQOLDALVKIVVARPETHMNEISDQLPISLSVA 5074
Qy 91 ----- 90
Db 5075 NSEPOSFVYKAGLSLVENHALNNSGGIALVFAHDIREGSRMESLTYGELNTRANQLAN 5134
Qy 91 -----DO----- 92
Db 5135 YLISQAKRDELICVMEKSVSLYLILAAVKAGCGYLPVPETPAARIRQILAEADVCF 5194
Qy 93 ----- 92
Db 5195 CLTDSSMAPVIADVSRCHEMNMVDTTDCSAQSGTGPQLDFKPTDIAAYVFTSGTTGPKGV 5254
Qy 93 ----- 92
Db 5255 LVTQENILSNLEVLKIVPVEGSRLLQACNOAFVDSVFEIFFTWYTCMLCSAKDVNF 5314
Qy 93 -----TDPGH----- 97
Db 5315 RDFEKAINELEITHLSLPTVAALTDPAHI PRVKELVTAGEAVTHVHGAWAGKLYQGY 5374
Qy 98 ----- 97
Db 5375 GPSETTICTVNSAVESDHVINNIGPAFENTSFAVLTOGDDFQLVPLGGLGELCFGGQOV 5434
Qy 98 ----- 97
Db 5435 FRGYQNPPELTSKINHNPNYGRIVRSGLRLLPDGTLIOGRTDDQDKIRGQRIELGE 5494
Qy 98 ----- 97
Db 5495 ISGCLLOFPSVQCAIEVIKTADKERLMAFWIPSGYSKDSYSILOPKNLEBIIKSIYAH 5554
Qy 98 ----- 97
Db 5555 LADNLPAYMVPDALVPVSAIPOTSQKIDKRLASDGSALTVEDLNAYSRGADDETSEL 5614
Qy 98 ----- 97
Db 5615 SATEOOLASALADTLOMSOTSIGRSTSPFALGLDSVSAIRLATNLKKEYCYSIDVSQILK 5674
Qy 98 ----- 97
Db 5675 RPTIARLPLLGESSKOTNEPVTADCEAAVGSYLHDESWSQLHGHGOTVSQLPCTPL 5734
Qy 98 ----- 97
Db 5735 QEAMLSARDTSGSAYRNKTLFSLHGSVDKLKACWEVMLQRHDLRTIFLSTEDSRFPVP 5794
Qy 98 ----- 97
Db 5795 QAVLSQWTLPWQECDDIPDQLTLLDSAKAGGDSIVDHSPPMKIQVYRSESTVYLLLDHM 5854
Qy 98 ----- 97
Db 5855 HALYDANAMSNLLYEQLYKQDSLAPVSKPFLNFMISTSVEEADALFRDQLREFVPK 5914
Qy 98 ----- 97

Db 5915 PFKRTDVKSGFGTITGRNLNYSKMWETFLSKHSTTMLSLTOAMMMKTLAASOSYSDVCCG 5974
Qy 98 ----- 97
Db 5975 NVVSGRSPVDGIESLVAPCFNTIPVRVLSKHSNGLVLCALQORVNIIDSLPYQLTPLRR 6034
Qy 98 ----- 97
Db 6035 IOAQAGTNGKRLPDSLVLLODDTTDLDSAIWRLEGESGVMDMYTSGTGLPKAVCLSHRA 6094
Qy 98 ----- 97
Db 6095 VTOSLLAHRFIPFSRFLQFASPTFDVSVEIFFPWRGATLVSVERNRLGLDLPCTIT 6154
Qy 98 ----- 97
Db 6155 SLNIDAAELTPSVAASLVRRHENVPTLRALLTIGEMLNTQVIOQFGSGICKSGLYGMYG 6214
Qy 98 ----- 97
Db 6215 PTEAAIHCTLOPGVDLPAGTIGIPLDTVSCFIVKPTSTKHASQLEILPIGEIGELVI 6274
Qy 98 ----- 97
Db 6275 GGHQADGYLNREEQTRAAFTHPKFGGLYRTGDKARLHRNGTLECYGRISSGOVKLRGQ 6334
Qy 98 -----A 98
Db 6335 RVELGEIEHAASKAGGCHAVIASVISGLLVLCIGDPHRVSSKDIKSAKOWLPAYMIPS 6394
Qy 99 DLVLY-----ITRF----- 107
Db 6395 DIVLLDDFPYLPCKVDKRLKLETDYNSNTAQHVSGSSDLSENAREITRIIESVLGVSIDH 6454
Qy 108 ----- 107
Db 6455 STDLSAAGLDSLRAIQVASQLRRQCADLGALELLSVSNVLDELVRKADENINNDND 6514
Qy 108 ----- 107
Db 6515 SEKNQOTVHELRSVERDFESKAFVSGIEDVLPCTPLODAMLVETAKRPOAYCNELRLTV 6574
Qy 108 ----- 107
Db 6575 SPKIPVERVROALPALAQRHTALRSFGMPSGVSHCAVTOVINKTLVTISQFAHVKSFTGW 6634
Qy 108 -----DLB----- 110
Db 6635 SVTNRETLRLPLHFQYKCSGAEBILVAIHHALYDQWSVEILEDLETLLEWERTPERPS 6694
Qy 111 ----- 110
Db 6695 FGAVNKFNLRRSBDQSHLDWFGEYLSVTPGRLPNLSPKMPPOPLQSIQHTIEMDM 6754
Qy 111 ----- 110
Db 6755 TLQAAHSYSCSAHVFFQAAVAILLGFYMGTEDTVFGTVFSGRTLPIVEISMVGPLLST 6814
Qy 111 ----- 110
Db 6815 LPTRINTLESRKFSVDLSRLQEDNRKIMRHSMTSLADIKKACGFNPGEAVFDSIFVMOET 6874
Qy 111 ----- 110
Db 6875 ARPDARAOTLLNLVEAHDYLEFNLTLEPTQGVKTKATYQSSLLPLQHWKTLQOLD 6934
Qy 111 ----- 110
Db 6935 LVKIVARPETHMNEISDQLPISVLSVANSEPSQFVYKAGLSLVENHALNNSGGLALVF 6994
Qy 111 ----- 110

Db 6995 AHDIRECTSRMESLTYGELNTRANQLANYLISQAKRDELICVMEKSVSLYLILAAVK 7054
Qy 111 ----- 110
Db 7055 AGCGYLPVPETPAAIRQILAEADVKFCLTDSMAPVIADVSRCHIMNVDTTDCSAQSC 7114
Qy 111 ----- 110
Db 7115 TGPQLDFKPTDIAYAVFTSGTCKPKGVLTQBNILSNLEVLKIYVPEGSRLQACNQ 7174
Qy 111 ----- 110
Db 7175 AFDVSVFEIFTWTGMCCLCSKDVWFRDFEKAINELEITHLSLPTVAALTDPAHIR 7234
Qy 111 ----- 110
Db 7235 VKFLVTAGEAVTHVHGANAGKGLYQGYGPSETTICTVNSAVESDHVINNICPAFENTS 7294
Qy 111 ----- 110
Db 7295 AFVLTQDDDFQLVPLGLGELCFGGQVFRGYQNMPELTESKIINHPNYGRYRSGLGR 7354
Qy 111 -LPDG-----NRVRG----- 120
|||
|||
Db 7355 LLPDGTILIQRTDQDKIRGORIELGEISGCLLOFPFVQNCIAEVIKTADKERLMFMI 7414
Qy 121 ----- 120
Db 7415 PSGYSKDSYILOPDKNLEBIIKSIYAHADNLPAYMVPALVPSAIPOTSOGKIDKR 7474
Qy 121 ----- 120
Db 7475 LASDGSALTVEDLNAYSGRADDSETSLSATEQOLASALADTLQMSQTSIGRSTFFALG 7534
Qy 121 -VTQ-----VTQ-----LOG- 126
|||
|||
Db 7535 LDSVSAIRLATNLRKEYGYSIDVSQILKPTIARLAPLLGGESSKOTNEPVTADCEAAVG 7594
Qy 127 ----- 126
Db 7595 SYLHDESVSQLEHGHGTQSVQVLPCTPLQEAMLSARDTSGSSAYRNKTLFSLHGSVDKX 7654
Qy 127 AC----- 128
||
Db 7655 ACHEVMQRLHDILRTIFLSTEDSRFPFVQAVLSQWTLPMQECDDIPQLSTLLDSAKAGG 7714
Qy 129 -SPTW----- 132
|||
Db 7715 DSIVDHSPPNKIQYRSESTVYLLDMHHALYDANAMSNLLYEVEQLYKQSLSPVSEK 7774
Qy 133 ----- 132
Db 7775 PFLNFMISTVEADALFRDLREFVPKPKRTDVKSGFGTITGRNLNYSPPMVFELSKH 7834
Qy 133 ----- 132
Db 7835 STTMLSLTOAMWMTLAASQSDVCCGVNVSGRSPVDGIESLVAPCFNTIPVRVDLSK 7894
Qy 133 ----- 132
Db 7895 HRSNLGLVKALQRVNIDSLPYQLTFLRRIOAAGTNGKRLFDSILVLLQDQDTLLDSAIR 7954
Qy 133 -----SCL----- 135
|||
Db 7955 LEGESGVDFPCIVELAPTNESTYLSLHFNRSYLDDVEVSNLHOACLSAFASCIRYPSD 8014
Qy 136 ----- 135
Db 8015 VSDFIDPADLVAGVLPDTHMQPVEAAKTRSEKSSGDSWSPLELQIRAAYSAVS 8074
Qy 136 ----- 135
Db 8075 SAPEDIRRDTTIYKLGUDSISAIQANRLRKDGLLVQASDVMEPSPCSELASAVQSRSQ 8134

Qy 136 --ITEDTQFD----- 143
: : :
Db 8135 TPVLDERGFDFEGFKHYRGAALQSHRIATEKVASVRECTPLQSGMLSEYTHSDGHQYFN 8194
Qy 144 ----- 143
Db 8195 HTFYAIEADIDSSKLOSAMSKVLEQHELLRTGTVGTTDDHEHPFVMLTYTEFVIDLETOA 8254
Qy 144 ----- 143
Db 8255 SREGSVVEYSEQKASESVKNNLHLPWRRWSLLGVGRCQLQFSAHHAIFDAESURLIMT 8314
Qy 144 ----- 143
Db 8315 DLQSALSNGYVPTRLTIDGALGHILNSQADVESORTFWSQKLSGAPVTRFPNMTFVRIS 8374
Qy 144 ----- 143
Db 8375 DTEANVELVLNKRKSLERCOBLGVSMSQVGOANARLLSAYTGESQVTFGVVLSGRT 8434
Qy 144 ----- 143
Db 8435 SPATADAAFPICITTLPVSTNTAVDDSQFLKOLMSYNATIOKHQFTPLTNIIRNYAESTSEA 8494
Qy 144 ----- 143
Db 8495 LFDLSFVYQRPMDVVDVSSWKIIREKASVELAVSMEALSEDGLGLRLTIDPAQVPE 8554
Qy 144 ----- 143
Db 8555 QGKIMLQOMVMIAGLLKPEDAINTSVMSIIPPKOPIIATDPKYLHEMTEASVKSYSORI 8614
Qy 144 ----- 143
Db 8615 AMEFVDALEDGOISSRHWTYRQLDDEANKIAHLLIDRGVKPGDIIATSFCKPEASFAFY 8674
Qy 144 ----- 143
Db 8675 GILKAGCAFDPTAPAARKAFILEDNSARVLLTSDSIRSELRELTOCDDIILINFENK 8734
Qy 144 ----- 143
Db 8735 NELSTSSVPVSGLPSSVSVLYTSGTTGPKCEITHDNVQVMSFKRLPKGRWTDSE 8794
Qy 144 ----- 143
Db 8795 RWLQFASVHFDVSVLEQFVTHVGMRLVCAPRDLILEDIAGFLDTMQITHLDLTPLSLRL 8854
Qy 144 ----- 143
Db 8855 LDPALVPSLHKGVFITGGBSLKQDQINTWGDVGLFNFGPTTECTIGTVTFPCVPKEGKP 8914
Qy 144 ----- 143
Db 8915 SNIGMDFNVGCYVLAPGTQTPVLRGAIGELCISGKLVGKYNRPBELTADCFPYLDARF 8974
Qy 144 ----- 143
Db 8975 ERYVRTGDLVRLPHDGSIDFLGRKDNQVKGRLRGORLEIDIEIAVIKRCODIQDTVCIVAKH 9034
Qy 144 ----- 143
Db 9035 PKQDKQLIAFIAGINESRKQKPELCPAESTRHLIOTARAACEERLPYVMVPTHFLPIOR 9094
Qy 144 ----- 143
Db 9095 IPLSVNNKVEKLLRQLYADLPTTVIQTATQADSOOSUSDEQKVAQALAEKLLKDDND 9154
Qy 144 ----- 143
Db 9155 LTPSSNIFSLGLSSISAIQFSKKLKASGFTTVQVATVKNPTISRLTKALATSTGRSGGE 9214

Qy 144 ----- 143
Db 9215 IADAKOVISACRQRHMGTVTRVLRCKADDIEAIPCTPFOGIIISRLASESSLYFNFSK 9274
Qy 144 ----- 143
Db 9275 FNAQGVLDQKLEGAFNQALERTQILRTFFIETDDGYYQAVRKGTGLPMTLEVVYDLASVD 9334
Qy 144 ----- 143
Db 9335 GVFAKRRQKWRSYNTSHLTVPFEIVVRSGSETFVSDLHLYDGNSTFDILMNNVSKLY 9394
Qy 144 ----- 143
Db 9395 NSQADFGKPFVDCIAGFPLRNQGAQKQFWDHLDLPDKVKSAMPPLIDNPASHDVLCTASL 9454
Qy 144 ----- LGVTI 148
Db 9455 DILNQADELRSLGTV 9471
RESULT 4
T30225
polyketide synthase - Streptomyces hygroscopicus
C:Species: Streptomyces hygroscopicus
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C:Accession: T30225
R:Apuricio, J.F.; Molinar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Khaw, L.E.; Staun
Gene 169, 9-16, 1996
A>Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hyg
A:Reference number: 220782; MUID:96186896; PMID:8635756
A:Accession: T30225
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-10223 <APA>
A:Cross-references: EMBL:X86780; NID:g987088; PID:g987099; PIDN:CAA60459.1
C:Genetics:
A:Gene: rapB
C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
C:Keywords: carrier protein
F:54-449/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F:1533-1604/Domain: acyl carrier protein homology <ACPl>
F:1647-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F:2135-2409/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F:3172-3243/Domain: acyl carrier protein homology <ACP2>
F:3287-3681/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F:3778-4052/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F:5160-5231/Domain: acyl carrier protein homology <ACP3>
F:5275-5667/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F:6776-6847/Domain: acyl carrier protein homology <ACP4>
F:6891-7285/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS5>
F:8411-8482/Domain: acyl carrier protein homology <ACP5>
F:8526-8921/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS6>
F:9012-9285/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F:10069-10140/Domain: acyl carrier protein homology <ACP>
Query Match 30.8%; Score 238; DB 2; Length 10223;
Best Local Similarity 1.0%; Pred. No. 92;
Matches 104; Conservative 17; Mismatches 25; Indels 9871; Gaps 32;
Qy 1 AAG----- 3
Db 86 AAGKSYCVGGFLDSAGGDFASFFGISPREALMDPQRLVLEASWEAFERAGIEPGLR 145
Qy 4 ----- 3
Db 146 GSDTGVMFCAYPGGYCVGADLGGFCGATAGAVSVLSGRVSYFFGLEGPVAVTDTACSSSLV 205
Qy 4 ----- 3
Db 206 ALHQAAALRQGECSLALVGGVTVMATPOSFVEFSRQRGLASDGRCKAFADSADGTGMAE 265
Qy 4 --GILHLELL-----VAV-----GP----- 16

Db 266 GAGVLLVERLSDAQKHQVLAIVRGSAVNODGASNGLSAPNGPSQORVIRAALSNAGLS 325
Qy 17 ----DVFOAH----- 22
Db 326 TAEVDVVEAHGTTGLDPIEAQALLATYGDREQLLLGSVKSLNIGHTQAAAGVSGVIK 385
Qy 23 ----- 22
Db 386 MMALQHLVPRTLHVDEPSRRHVDWTDGAVALVTENQPMPCRRPRAGVSSFGISGTNA 445
Qy 23 ----- 22
Db 446 HVILESAPPTQAVDDVPPEAPVWASELVPLVISARTLPALVEYEDRLRAYLAASPGVDV 505
Qy 23 ----- 22
Db 506 RGVASTLAVTRSVFEHRAVLLGDGDTVTGTVSDPRVVFVPPGQSORAGHGEELAAFPV 565
Qy 23 ----- 22
Db 566 FARIHQQVWGLLDVPLEVNETGYAQPALFALQVALFGLLESWGRVDPDAVVGHVSGELAA 625
Qy 23 ----- QEDTERYVL----- 31
Db 626 GYVSGLSLEDACITLV SARARLMQALPPGGVMVAVPVSEDEARAVLGEVGEIAAVNGPSS 685
Qy 32 ----- 31
Db 686 VILSGDETAVLQAAALGKSTRLATSHAFHARMEPMLEEFRTVAERLTYTQTPRLAAG 745
Qy 32 ----- 31
Db 746 DRVTTAEYVRQVDRTVRFGESQVASYEDAVFIELGADRSRLARLVGVMMLHTDHEAQAAI 805
Qy 32 ----- 31
Db 806 SALAHLVNGVTVDWTALLGDAPATRVLDLPTVAFQHORYWLEGADRAAGHPLLGPAV 865
Qy 32 ----- 31
Db 866 GLAASGVLFTSQVSRSGDLWLQDOTVLPATAPAEMLAALAADETCGLVEELTVEALLLL 925
Qy 32 ----- 31
Db 926 PDGAVQVQVWSEPPDDAGRRLSIHARYSDELWTRLATATLATTGTVSGWQAGEAWP 985
Qy 32 ----- 31
Db 986 PAGAVPVEARVPSLRGVWRGSEVFAEVALDDTHDTRFLLHPGLMTAARMAVGEETPAA 1045
Qy 32 ----- 31
Db 1046 WQGLTHARNAELRVRLTSODDETLSVEAVDSTGLPVLARSALDAVPVNEPATND 1105
Qy 32 ----- TNLNICA----- 38
Db 1106 LLTLTWAGIATROOQTGLTGAFEDLAADGAPVDPVWVFTALPGNDDPLAQTRILTAQV 1165
Qy 39 ----- 38
Db 1166 LRTVOEWIGGERFSDSTLVLRVTGTGLAAAVSGLMRSAQSEHPGRFILVESDDDDALTPDQ 1225
Qy 39 ----- 38
Db 1226 LAATVGLDEPLRLISDGRYEVPRLTRAHADPEPERAWDPDGTVLTITGSGVLAGIVARH 1285
Qy 39 ----- 38
Db 1286 LVAERGVRHLLLSRGTPDRALLSELAEALGAAVDTACVSDRAELARVLARVSPEHPLT 1345
Qy 39 ----- 38

Db 1346 AVIHTAGVDDGVVESLSAORLETVPFRPKADGAWHLHELTRDADLAAAFVMSAAAGUMAG 1405
QY 39 ----- 38
Db 1406 AGOGNYAANAFLDALAEERRAEGLPALAVAWGLMEDASGLTAQLTDTDRDIRRGGLRA 1465
QY 39 ----- 38
Db 1466 ISAEHGMRLPDNASRHSEPLVAAPMEPVRDAEVPALLRSLHRPNVRRRAALAGGAQWLAA 1525
QY 39 ----- 38
Db 1526 LAPEERAKALLKVVRDTAATVLGHADARTIPVTGAFRDLGIDSLTAVELRNLAKVTGLR 1585
QY 39 ----- 38
Db 1586 LPATLVFDYPTPAVLAARLCELFTGENPVLVRTASVVGODEPLAIVGMACRLPGGVSSPE 1645
QY 39 ----- 38
Db 1646 DLWRLVESGTAISGFPADRGWDAESLFPDPAVGKSYCEGGFLDSAAASFDAGFFGIS 1705
QY 39 ----- 38
Db 1706 PREALAMPDPOORLIMEVSWEAERAGIEPGSVRGSDTGVFMGAYAGGYGAGADLGGFAAT 1765
QY 39 ----- 38
Db 1766 ASATSVLSGRVSYFFGLEGPATVDTACSSSLVALYQAGYALRQGECSLALVGGVTVMAT 1825
QY 39 ----- 38
Db 1826 PQSFVEFSRKSGLSDGRCFAFADSADGTGWAEGVGVLLVERLSDAQAKGHQVLAVRSS 1885
QY 39 ----- 38
Db 1886 AVNQDGASGLSAPNGPSQGVIRQALANAGLTTAEVDVVVEAHGTGTLGDIPEAQAVIA 1945
QY 39 ----- 38
Db 1946 TYGODRERPLLLGSLKSNIGHAQAASGVSVIKVMALQHNTVPTLHVDEPSRHVDWAA 2005
QY 39 ----- 38
Db 2006 GAVELVRENQWPCTDRPRAGVSSFGVSGTNAHILESAPPAQAEAAQVETPPVVASD 2065
QY 39 ----- 38
Db 2066 VPLVISAKTPALTEHEDRLRAYLAASPGVDTRAVASTLAVTRSVFEHRAVLLGDDTVT 2125
QY 39 ----- 38
Db 2126 GTAVSDPRVVFPFGQGWLGMGSALRDSIVFAERMAECAPALREFVDWDLFTVLDDP 2185
QY 39 ----- 38
Db 2186 AVVDRVDVQPASRMVSLAAVMQAAGVRPDAVIGHSQGEIAAACVAGAVSMRDAARIV 2245
QY 39 ----- 38
Db 2246 TLRSEAIARGLAGRAMASVALPAQDVELVDGAWIAAHNGPASTVIAGTPEAVDHLTAH 2305
QY 39 ----- 43
Db 2306 EARGVRVRITVDYASHTPHVELIRDELDTITSSSSQAPVVPWLSTVDGSWDSPLDVE 2365
QY 44 ----- 43
Db 2366 WYRNLREPUGFHPAVGOLQAEQDTVFEVSASPVLLQAMDDDDVVTVATLRRDDGDATRM 2425
QY 44 ----- 47
Db 2426 LTALAAQAVHGVTDWPAILLGTATTRVLDTLPTAFQHORYWLRSDRAAADGHPLLGTVV 2485

QY 48 ----- 47
Db 2486 DLPASDGVVLTGRVSLATHTWLADHAVRGSVLLPGTAFVDLVVRAADEVESDVVDELVIE 2545
QY 48 ----- 47
Db 2546 TPLLLLPQTGGVQLSVSVGGADESGHRAVMVFSQADNTDTWTRHTVATVSTSDTSVSLPEF 2605
QY 48 ----- AQ 49
Db 2606 ASWPPAQAARPVSVADFYDRLAAAGTEYGPAGFQGLQAAMRGDVTVAEVVLAEBOAQEAAR 2665
QY 50 FRVH----- 53
Db 2666 FAVHPALLDAAMHASVLHTPDTDOOSVRMPFSWNHVQIRATDTAMLURVAATPTTDGWSVR 2725
QY 54 ----- 53
Db 2726 VADDTGRPVATTIGSLVTRPVVTADTLGSAADDLLTLVWTEIPTPQSSLSVGRYEDLADGD 2785
QY 54 ----- 53
Db 2786 VPPVDVAVYTARPEADGSPDLVQTRTLTAQVLQTAQAWLGGERTDSTLWVRTGTGLAA 2845
QY 54 ----- 53
Db 2846 AAVSGLMRSAQSEHPGRFVLVESDDSLTLHQAATVGLDEPRLRVNDGRFEVPRLARVN 2905
QY 54 ----- 53
Db 2906 AAEPSETVNDPHGTVLTITGSGVLAGALRHUVTERGVRRHLLLSRTTAD EGLNELGE 2965
QY 54 ----- 53
Db 2966 LGARVETADCDVSDRAGLARVLAVGVSPEHPLTAVIQTACAGLDGVLTLTAQRDLTVLRP 3025
QY 54 ----- 53
Db 3026 KADGAWHLHELTRNTGLAAAFVMSYSAAGVMPGQGNLAAATAFLDALADQRRAEGLPAL 3085
QY 54 ----- 53
Db 3086 ALAWSSETSDLIGLRTISABRGMRLFDSASHRGEPPLMAASLDPARAAEVPAALLRSLR 3145
QY 54 ----- 53
Db 3146 RPVARRAASADGGVQWLAALAPAEREKALLKVVCDSSAAVVULGHADARTIPVTGAFKDLGV 3205
QY 54 ----- 53
Db 3206 DSLTAVELRNLVKATGLRLPATMVFDYPTTALAARLDELFTGENPAPVREPVPVAQD 3265
QY 54 ----- 53
Db 3266 EPLAIVGMACRUPGGVSSPEDLWRLVESGTDVSGPFTDRGWDVEGLFDPDPAACKSYR 3325
QY 54 ----- 53
Db 3326 AEGGFLLDTAAGFDAGFFGISPREALAMDQQRLLLLEVSWEAFERAGIEPGSVRGSDTVGF 3385
QY 54 ----- 53
Db 3386 IGAFPPVYGAGAAREGYGATAAPNVLSGRLSYFFGLEGPATIMDTACSSSLVALHLAAQA 3445
QY 54 ----- 53
Db 3446 LRNGCSMALAGGVYVMATPEVTFEPARQGLASDGRCKAFADSADGAGFSEGAGLLLVE 3505
QY 54 ----- 53
Db 3506 RUSDARRNGHQLAVVRGSAVNODGASNGFTAPNGPAQORVIRQALANAGLTTAEVDVVE 3565


```
QY 54 -----LVKWI----- 59
Db 3566 AHGTGTTLGDPIEAQAVIATYQDREQPLLLGLTKSNVGHQTQAAAGVSGVIMWALQHS 3625
QY 60 ----- 59
Db 3626 TVPRTLHVNEPSRHYDWSAGAVELTENQSMVPTGPRRAGVSAGVSGTNAHVLESAP 3685
QY 60 -----LPE----- 64
Db 3686 PAQSNNAPVATPVVASLPLVISAKTLPALTEHEDRLRAYLAASPCADMRAVGSTLA 3745
QY 65 ----- 64
Db 3746 LTRSVFEHRAVLGHDTVTCTGTAVSNPRVVFVPCGQWGLMGSAALRGSSVVFPAER 3805
QY 65 ----- 64
Db 3806 MAECAALSEFVDWDLFAVLDDPAVVDRVDVVQPASWAMVSLAAVQADGVRPDAVIGH 3865
QY 65 ----- 64
Db 3866 SOGEIAAACVAGASLRDAARSVTLRSQAIARGLAGAMASVALPAHEIELVDGAWIAA 3925
QY 65 -----GAP-----NITANLTS 75
Db 3926 HNGPASTVVAGAPEAVDRVLAVHEARGVRRIANDYASHTPHVELIRDELDDITAGIGS 3985
QY 76 SL----- 77
Db 3986 QAPVVPWLSTVDTWVEGLDVEYMYRNLRPEVPGFDSAVQLRAEGDVFVEVSASPVL 4045
QY 78 ----- 77
Db 4046 QAMDDDVVTVATLRRDDGATRLMTALAQAQFVEGVTVDPWPAILGATTRVPDLPYAFQH 4105
QY 78 ----- 77
Db 4106 QRFWAEGADRSVAGGHPLLGVAVELPESDGVLLTGRVSLATHANLADHAVRGSVLLPGTG 4165
QY 78 ----- 77
Db 4166 FVELVRAADEVCGDVDELVIETPLLLPSSASVHLSVSGEADSGRGRVTVFSRADGA 4225
QY 78 ----- 77
Db 4226 DAMTRHVSATIGVGAALSILPELAAMPQAQPVGLGFYDRLTGAGYEGPAFQGLQAA 4285
QY 78 ----- 77
Db 4286 WRDGDTVFAEVALAEQAEAAAFVHPALLDAALDAGILNTLDNAEQGVRLPFSWNGVQ 4345
QY 78 ----- 77
Db 4346 VRATGSATLRVADNPTSDGWSVRVADDSGRPVATVDSLVRPVTADTLGSAADLLTVW 4405
QY 78 ----- 77
Db 4406 TEIPTQOTGLSVGRFEDLADGVPVPEVVVCTALPDSENPLAPLDPDPLVQTRTLTT 4465
QY 78 ----- 77
Db 4466 QVLQAOVWLAGERFDTSLVVRTGTGLATAGVSGLMRSQAQSEHPGRFVLVECDNLTLO 4525
QY 78 -----LSVC----- 81
Db 4526 QLAATVGLDEPRLEVCGRFEVRLARANTPESSPLTIPCDRAWLLEQSHSGTLRLDALV 4585
QY 82 ----- 81
Db 4586 PAETAPERLOSQGEVRVDVRAAGLNFDRDLIALGTYPGEAIVGAAGVWLEVGPEVQDLA 4645
QY 82 -----GWSQT----- 86
```

```
Db 4646 PGDRVFELVGGFGGFAVAIADRRMLGVIPOGWSFTTAASVPVVFATAYYGLVDLAGLSAGE 4705
QY 87 ----- 86
Db 4706 SVLIHAAAGVGWMAQTQIARHLGARIYATASTGKHVLEAGLEDARIGDSRTTGPREM 4765
QY 87 ----- 86
Db 4766 LDTTDSRGVDVNLNSLGDGFDASDLPLPRGGRFVEMGKTDIRDPHQVTAADRPQTSYQAF 4825
QY 87 ----- 86
Db 4826 DLMDAGPDLREIITADLLALFAQGVLLPLPVRAWDIRQAREAFSWMRSRARIHGIKIVLTVP 4885
QY 87 ----- 86
Db 4886 QQLDANGTVLTGSGVGLAGIAARHLVAEOGVRRHLLLSRSTPDDALINELGELGARVDT 4945
QY 87 -----INPE----- 90
Db 4946 AICDVSDRAGLARILAGVSPHEPLTAVIHTAGALDDGVWESLTAQQLTVLRPKADGAWH 5005
QY 91 ----- 90
Db 5006 LHELTRDADLAAAFVWYSSAAGVLGSGGQGNYYAANAFDLAEORRGEGLPALAVAGLW 5065
QY 91 ----- 90
Db 5066 EDASGLTAEMTDDTRDIRRGGLRAISAGHGMGLLDAARHGEPVLLAAAMEPVREAEVP 5125
QY 91 ----- 90
Db 5126 ALLRLLHRPVARRAASTGSSVQWMLARLAPVEREKALLKLVCDGAATVLGHADASTIPAT 5185
QY 91 ----- 90
Db 5186 AAFKDLGIDSLTAVELRNSLTAKTGLRPLATLPDPTPTALAAALGWFVGETPVPVRT 5245
QY 91 ----- 90
Db 5246 SVSVVAQDEPLAIVGMACRLPGGVSSPEDLWRLESSTDVSGPPTDRGWDVNLFGPAA 5305
QY 91 ----- 90
Db 5306 GDSYRLQGGFLDAAAGPDASFFGISPREALAMDQQRLVLEVSWEAFERAGIEPGSVRGT 5365
QY 91 ----- 90
Db 5366 DTGVFMGAYPGGYGIGADLGGFGATASAVSVLSGRVSFFGLEGPAITVLMACSSSLVAL 5425
QY 91 ----- 90
Db 5426 HQAGYALRQGESLALVGGVTVMATPQTFEFARQGGLAGDGRSKAFADSDAGAGFSEGV 5485
QY 91 ----- 90
Db 5486 GVLLVERLSDAQAKHQHVLAMLRSSAVNQDGNLTPNGPSQORVIOAALSNAGLAH 5545
QY 91 ----- 90
Db 5546 EVDVVEAHGTGTTLGDPIEAQALLATYQDREQPLLLGSKVSNLGHQTOAAAGVSGVIKHW 5605
QY 91 ----- 90
Db 5606 MALQRFVPRTLHVDEPSRHVMSAGAVALTENQPHDMPGRARRAGVSSFGISGTNAHV 5665
QY 91 ----- 90
Db 5666 ILESAPPTQPADNAVIERAPEWLPWVISARTQSALTEHEGLRAYLAASPGVDMRAVAST 5725
QY 91 ----- 90
```

Db 5726 LAITRSVFHRAVLLGDDTVTGTAAATDPRVVFVPPGQSGRAGMGEELAAAFVFAIRHQ 5785
Qy 91 ----- 90
Db 5786 QVNDLDDVPDEVNETGYAQPALFALQVALFGLLESWGVPRDAVVHSGVSELAAAGYVUGL 5845
Qy 91 ----- 90
Db 5846 WSLLEDACTLV SARARLMQALPAGGVMAVVPVSEDEARAVLGEVGEIAAANGPSSVVLSGD 5905
Qy 91 ----- 90
Db 5906 EAAVLOAAEGLGWKTRLATSHAFSARMEPMBLEEFRAVAEGLTYRTPQVMAAGDQTTTT 5965
Qy 91 ----- 90
Db 5966 EYVVRQVRDTRFGEQVASYEDAVFVGLGADRSRLARLVGVAMLHGDHEAQAASALAH 6025
Qy 91 ----- 90
Db 6026 YVNGVTVDMPALLGDAPATRVLDLPTYAFOHRYWLEGTDRMAAGHPHLLGEAVAVPGSD 6085
Qy 91 ----- 90
Db 6086 GVLLTGRVSLATHPWLADHVRGVSLLPGTGFVELVVRAADEVGCDVIDELVIETPLLP 6145
Qy 91 ----- 90
Db 6146 QTGCVQLSVAVESDDSGRRAVTVFSRADNVDTWRHATVTVNSDITLSPDLASWPPA 6205
Qy 91 ----- 90
Db 6206 QAQPVNVADFYDQLTVVAYEYGPAPFQGLQAARWGDVTVAEVALAEQAQDAARNAVHPA 6265
Qy 91 ----- 90
Db 6266 LLDAALHAGILNTPDTEQGVRLPESMNQVQVHAVGSAMLRVAVTQAADGHSVRVADTGR 6325
Qy 91 ----- 90
Db 6326 PVATIGSLVTRPVTAOTLGSTDDLLALTWTEIPAQOETSLTIGRPFEDLADADVPEVV 6385
Qy 91 ----- 90
Db 6386 VFTARPDSSENPLAQTRTLTAQVLOKIQAMLABERFTDSLIVRIGTGLAAAVSGLMRS 6445
Qy 91 ----- DDT ----- 93
Db 6446 QOSEHPCRFVLVESDDTLAPDOLAATVELDEPLRLISDGRFVPRLTRTHAAEPESEKV 6505
Qy 94 ----- 93
Db 6506 WDPDGTVLITGGVLAGIAARHLVAERGVRRHLLLSRSAPDGGALISELGELOAVATV 6565
Qy 94 ----- 93
Db 6566 CDVSDRGLVRVLADVSPHPLTAVIHTAGVDDGVVESLTAQRLDVTLRPKADGAMHLH 6625
Qy 94 ----- 93
Db 6626 ELTRDADLAAPVMSAAGVFGSAGQNYAVANFLDALAEQRRRAEGLPALALAWLWEG 6685
Qy 94 ----- DP ----- 95
Db 6686 TSGTLANTDTHDRIRRSGRMAISAEBGMRLFDGASRRRDPVLVAAMEPVEAEVPA 6745
Qy 96 ----- GHADL ----- 100
Db 6746 LRSLRPVARRAASGAVRCLAALAPERAKALVKVVCDSAAATVLGHADVDSIPVTAFR 6805
Qy 101 ----- IIII ----- 100
Db 6806 DLGVDSLTAVELRNSLTATGLRLPATLVFDYPTPGALAAARLEELFTGENPVQVTRPVSA 6865

Qy 101 ----- 100
Db 6866 VCODEPLAIVGMACRLPGCVSSPEDLWRLVSGTDAISGFTDRGWDVENLFDSDPDAAG 6925
Qy 101 ----- 100
Db 6926 KSYCVEGGFATAANFDASFFGISPREALAMPQORLVLEVSWEPERAGIBPGSVRGSD 6985
Qy 101 ----- 100
Db 6986 TGVFMCAFFGGYGIGADLEGYATAGLNVLSGRLSYFFGLEGPATVTDACSSSLVALHQ 7045
Qy 101 ----- V 101
Db 7046 AGVALRQGECSLALIGGVTMATPHTFVBFSRQGLASGRCKAFADSADGTGWSGV 7105
Qy 102 LYITRF ----- 107
Db 7106 LLVERLSDAQAKGHQVLAIVRSVAVNQDGSNGLSAPNSQORVIRQALANAGLTTAEV 7165
Qy 108 ----- 107
Db 7166 DVVEAHGTGTTLGDPIEAQAVIATYQDRDQPVLLGSVKSINVCHTQAAAAGVSGVIKVMMA 7225
Qy 108 ----- 107
Db 7226 LQHGLVPRTLHVDEPSRHYDWDGAVELVTENQSWPEACRPRRAGVSVFCVSGTNAHVIL 7285
Qy 108 ----- 107
Db 7286 ESAPPTQAVDDVRPADAPVVASVMASELPLVISAKTOSALAEYEGRLRAYLAASPGVDM 7345
Qy 108 ----- 107
Db 7346 RAVASTLAMTRSVFHRVAVIGDDTVSGTAATDPRVVFVPPGQSGRAGMGEELAAAFV 7405
Qy 108 ----- DLE ----- 110
Db 7406 FARIHQVWDLDDVPDEVNETGYAQPALFALQVALFGLLESWGVPRDAVHSGVSELAA 7465
Qy 111 ----- LPDGN ----- 115
Db 7466 AYVSGLSLEDACTLV SARARLMQALPAGGVMAVVPVSEDEARAVLGEVGEIAAANGPSS 7525
Qy 116 ----- 115
Db 7526 VVLSGDEAAVLOAAEGLGWKTRLATSHAFSARMEPMBLEEFRAVAOGLTYHAPGVVMAAG 7585
Qy 116 ----- ROVR ----- 119
Db 7586 DRVMTAEYVVRQVRDTRFGEQVASYEDAVFVGLGADRSRLARLVGVAMLHGDHETQAAI 7645
Qy 120 ----- GVT ----- 122
Db 7646 GALAHLVYNGVTVDMTALLGDVPVTRVLDLPTYAFOQRYMAEVRGSRADVSGAGLDAVGH 7705
Qy 123 ----- 122
Db 7706 PLLGAVVALPGSDGVLLTGRVSLATHAWLADHVRGVSLLPGTGFVELVVRAADEVGCDV 7765
Qy 123 ----- 122
Db 7766 VDELIVEAPLLLPQTGGVQVSVSVEADESGHRVVTIFSRADNADTWVRHVS AVISTDS 7825
Qy 123 ----- 122
Db 7826 TVSLSEFASWPPAQAPVDVTGYDPLMAEGYEGPAGLOAAMRYGDTVFAEVALAE 7885
Qy 123 ----- 122
Db 7886 QVREAAARYPSVHPALLDAAALHACTLECAAEVGVGLPFSWDGVRVHAGGSAMLRVAVTQA 7945

QY 123 ----- 122
 Db 7946 ADGWSVRVADEIGRPVANVSLVTRPVTDALGSAADLLALTWAEIPAPEGTLTVGRF 8005
 QY 123 ----- 122
 Db 8006 BELVSDGDPVPEVAVFTALPDNDPPLEQTRKLTGQVLQAVQEWLGGERSDSLTVVRT 8065
 QY 123 ----- 122
 Db 8066 GTGLAAAGVGLMRSQAQSEHPGRFVLVESDDDALALDQAAAAGLDEPRLRISDGRFEAP 8125
 QY 123 ----- 122
 Db 8126 RLTRTHAAPESEKVMKDPDGTVLJTGSGVLGAGIAVRHLVTERGVRHLLLSRSAPDEAL 8185
 QY 123 --OLG----- 125
 Db 8186 INQLGELGARVETAACDVSDDRAALAQVLGAGVSPHPLTAVIHTAGVLDGCVVESLTAQRL 8245
 QY 126 ----- 125
 Db 8246 DAVLRPKADGAWNLHELTRDADLAAAFVMSAAGVLSGGGQGYAAANAFDALAEORHA 8305
 QY 126 ----- 125
 Db 8306 EGLPALAVAWGLWEDASGLTAQLTDTDRIRRGGLRAISAHEGMGLFDSASRHSEPVIV 8365
 QY 126 ----- 125
 Db 8366 AAPMEPVRDAEVPALLRSLRHPVARRAAAGGARWLALAPAEREKALLKLVSDCAATVL 8425
 QY 126 ----- 125
 Db 8426 GHADTSTIPATTAFKDLGIDSLTAVELRNSLAKATELRLPATLVFDYPTPTALAAARLDEL 8485
 QY 126 ----- 125
 Db 8486 FTGENPAPVRGPVSAVAODEPLAIVGMACRLPGGVSSPEDLWRLVESGTDIAISGPTDRG 8545
 QY 126 ----- 125
 Db 8546 NDVENLYDPDPAPGKSYSQGFLDAAAGFDASFFGISPREALAMDPOORLMLVSWEA 8605
 QY 126 ----- 125
 Db 8606 FERAGIEPGSVRGSDTGVFIGAYPGYGIGADLGGFTTAGAASVLSGRVSYFFGLEGPA 8665
 QY 126 -----GACS-----PT----- 131
 Db 8666 FTVDTACSSSLVALHOAGYALRQGECSLALVGGVTVMPTPQTFVFSRORGLSADGRCKA 8725
 QY 132 ----- 131
 Db 8726 FADAADGTGAEGVGVLLVLERLSDAQANGHQILAVRSSAVNQDGNGLSAPNGPSQR 8785
 QY 132 ----- 131
 Db 8786 VIRAALSAGLAPHEVDVVEAHGTGTLGPIEAQAVIATYQGRGEPLLLGLSKSNVGH 8845
 QY 132 -----WS----- 133
 Db 8846 TQAAAGVSVIKVMALQHSMPRTLHVDEPSRHRVDSAGAVELVAENQFMPETGRPRRA 8905
 QY 134 ----- 133
 Db 8906 GVSSFGISGTNAHVILESAPASQVGTAGTPTVLVSELVPLVISAKTQPALTEHEDRLRA 8965
 QY 134 ----- 133
 Db 8966 YLAASPGVDIRAVASTLAVTRSVFEHRAVLGDETVTGTAVSDPRIVFVFPQGQMWLGM 9025
 QY 134 ----- 133

Db 9026 GSALRDSSVFAERMAECAAALSEFVDMDLFAVLDDPAVVDRVDVVQPASWVMVSLAAV 9085
 QY 134 -----CL----- 135
 Db 9086 WQAAGVRPDAVIGHISOGEIAAACVAGAVSMRDAARIIVTLRSQIARGLAGRGAMVASVALP 9145
 QY 136 ----- 135
 Db 9146 AOVVELVDGAWIAAHNGPASTVIAGTPEAVDHLVTALRQRGAGNADHDVYASHTPHVELI 9205
 QY 136 -----ITED----- 139
 Db 9206 RDELLDITSSSSQDPLVPLMLSTVDGTWVDSPLDGEVWYRNLREPVGPHPAVSOLQAQGD 9265
 QY 140 ----- 139
 Db 9266 TVFEVSASPVLMOAMDDDDVVTVATLRRDDGDATRMLTALAQAAYVHGVTVDMRAVLGDVP 9325
 QY 140 ----- 139
 Db 9326 ATRVLDLPTTAFQHORVWAEAGRSADVSAAGLDAVGHPLLGAVLAMPGSDGVMLTGRVSL 9385
 QY 140 -----TCF----- 142
 Db 9386 ATHAWLADHAVRGSVLLPGTGFVELVVRAADEVACDVVDELIVEAPLLLPOTGGVQLSVS 9445
 QY 143 ----- 142
 Db 9446 VGEADESCHRAVTVFSRADSADAWVRHVSATVSVSDTTVTPTSDLTAWPPQAOKPVDVAGF 9505
 QY 143 ----- 142
 Db 9506 YDQLTRAGYEGPAGFQGLQAARWGDVTFAEVALAEQOTQDAARFAVHPAVPAVLAALHAGI 9565
 QY 143 ----- 142
 Db 9566 LNTPDADRDTVRLPFSNMHVQVHVTSATLRVAMTRVADGWGVRVADDIGRPVATIGSLV 9625
 QY 143 ----- 142
 Db 9626 TRPVAADALGSAVDDLFALTWTEIPVSQQGVTVGKFEDLADGEVPMPPDVVFTALPDSG 9685
 QY 143 ----- 142
 Db 9686 DPLAQTRRLTAELVLOAVQVWLAGERTDSTLVVTRGTGLAAAAVSGLMRSAQSEHPGRFV 9745
 QY 143 ----- 142
 Db 9746 LVESDDDDTLTPDQLAATVGLDEPRLRVLDGRYEAPRLTRTGTVAEPEPEGVMDPDGTVLIT 9805
 QY 143 ----- 142
 Db 9806 GSGVLAGIAARHLVAERGVHLLLSRSAPDEALISELAEAGAAVDTAVCDVSDRAGL 9865
 QY 143 ----- 142
 Db 9866 ARVLGVSPPHPLTAVIHTAGVLDGCVVESLTARRLOTVLRPKADGAWNLHELTRDIDL 9925
 QY 143 ----- 142
 Db 9926 AFVMSYSSAAGVLSAGQGNAYANAFVDALAEORRAEGLPALALAWGLWEDASGLTAKLT 9985
 QY 143 ----- 142
 Db 9986 GTDHDRIIRSGLTITAERGMLFDIASROGEPVLVATPMEPVEVEVPALLRLLHRPVA 10045
 QY 143 -----DLGV 146
 Db 10046 RRAASTGSSAQNVLGLAPEERAKALLKVVTRDSATVGLHADARSIPATGAFKDLGV 10102
 RESULT 5

T14593
synergomycin synthetase - Pseudomonas syringae pv. syringae
C:Species: Pseudomonas syringae pv. syringae
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 01-Dec-2000
C:Accession: T14593
R:Guenzi, E.; Galli, G.; Grgurina, I.; Gross, D.C.; Grandi, G.
J. Biol. Chem. 273, 32857-32863, 1998
A:Title: Characterization of the synergomycin synthetase gene cluster. A link between ph
A:Reference number: Z18153; MUID:99047670; PMID:9830033
A:Accession: T14593
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-9376 <GUE>
A:Cross-references: EMBL:AF047828; NID:G3510628; PID:G3510629; PIDN:AAC80285.1
C:Genetics:
A:Gene: syrE
C:Superfamily: acetate-CoA ligase homology; acyl carrier protein homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:441-896/Domain: acetate-CoA ligase homology <ACL1>
F:914-981/Domain: acyl carrier protein homology <ACP1>
F:1529-1984/Domain: acetate-CoA ligase homology <ACL2>
F:2002-2089/Domain: acyl carrier protein homology <ACP2>
F:2613-3071/Domain: acetate-CoA ligase homology <ACL3>
F:3089-3156/Domain: acyl carrier protein homology <ACP3>
F:3700-4158/Domain: acetate-CoA ligase homology <ACL4>
F:4176-4244/Domain: acyl carrier protein homology <ACP4>
F:4768-5224/Domain: acetate-CoA ligase homology <ACL5>
F:5242-5310/Domain: acyl carrier protein homology <ACP5>
F:5834-6280/Domain: acetate-CoA ligase homology <ACL6>
F:6298-6366/Domain: acyl carrier protein homology <ACP6>
F:6892-7352/Domain: acetate-CoA ligase homology <ACL7>
F:7370-7437/Domain: acyl carrier protein homology <ACP7>
F:7990-8440/Domain: acetate-CoA ligase homology <ACL8>
F:8458-8526/Domain: acyl carrier protein homology <ACP8>
F:9015-9083/Domain: acyl carrier protein homology <ACP9>
F:946, 2034, 3121, 4208, 5274, 6330/Binding site: phosphopantetheine (Ser) (covalent) #status

Query Match 30.7%; Score 237; DB 2; Length 9376;
Best Local Similarity 1.2%; Pred. No. 76;
Matches 94; Conservative 24; Mismatches 28; Indels 7809; Gaps 29;

QY	1	AAGG-----	4
DB	980	AAGGSEVVVPAIPEHCDRITQMLPLINLSQAQIDHVVVKMPGGVANNQDIYPLAPL	1039
QY	5	-----	4
DB	1040	OAGILYHHISAEQDPTLKALFALSDRAQLDFSGALQGVINRHDLRTAVLWEDLDEP	1099
QY	5	-----	4
DB	1100	VQVVLQRELHVTELLDPDVGPDVEQLHQRFDRRHRLDVRTAPLMRIVFSDHPVNDRW	1159
QY	5	-----	4
DB	1160	LAMLLCHHLVSDATSLSVILOEIQAHLLGQGNALGEAVPYRNVYAQLRGVSEAQAFAFF	1219
QY	5	-----ILHL--	8
DB	1220	RDMLGDIDEPTLPFGLQDVQDGRNLEBASVTLAAELNRLRAQARQAGVSASLMHLAW	1279
QY	9	-----	8
DB	1280	ARVLGNVSACQVVGTVLLGRMQGADRALGMFINTLPLRVDIGATTVAEGLKATHE	1339
QY	9	-----	8
DB	1340	RLTALLGHEAPLVLAQRCSGVAAPTPLFSALLNRYHSVSDVRQKPLGLGSSLRHRP	1399
QY	9	-----	8
DB	1400	YQLPVDNLVDLGDAFTLTIAVQOISATRIGEYMQVALRNLDALHTPQALNSLSIL	1459

QY	9	-----	8
DB	1460	PDDERELLLTGNDTAHPYPRDVLIHOLIEQQAQRPDACA VRGDSGTLTLYAELNOQAN	1519
QY	9	-----ELLVA--	13
DB	1520	QLAHLRIELGVEDPTRVAVSLRGAEMVVALLGILKAGGAYVPIDPDLPARQAYMLSDS	1579
QY	14	-----	13
DB	1580	APRAVLTSHELLADLPDLGVPALVLDGRDSDALLKKQPTGNPDAKALDLPQNHLYVLYT	1639
QY	14	-----	13
DB	1640	SGSTGTPKGYMNEHLGVNRLWARDAYQVNSQDRVLTQKTPFGFDSVMVEFFPLLTGAE	1699
QY	14	-----	13
DB	1700	LVNARPSGHQDPYLAQVISDAGITLLHFVPSMLDVFLHRSTRDPQLRRVLCSGEALP	1759
QY	14	-----	13
DB	1760	RALQRFEQHLKGVELHNLGYPTAAIDVTAMECRPTDPCGDSVPICRPIANIQMHLDAL	1819
QY	14	-----	13
DB	1820	GQLQPMGVAGELHIGGIVARGVYNQPLSAERFIADPFSDNPQARLYKTGDVGRMLANG	1879
QY	14	-----	13
DB	1880	ALEYLGRNDFQVKIRGLRIEIGEIEAALAKHPAVEAVVTAREDI PGDKRLVAYYTQTA	1939
QY	14	-----	13
DB	1940	HTAVDLETLRGLHQVLPPEYMWPAIYVLLLEAMPLTNGKLDRAKALPAPGDALISRGYEA	1999
QY	14	-----	13
DB	2000	POGEIEQIAVIQDLLGVEQVGRHDHFPGLGHGSLAVSLIDRLRKQDLNLTNTVFTA	2059
QY	14	-----	13
DB	2060	PSVREMAIAISOEQALFRAPANRI PAHCTQLTPDMLPLVELSVAQIELIASAVPGGAN	2119
QY	14	-----	13
DB	2120	IQHIYPLAPLOEGILFHYLLNLERDAYLVYSTIEFDSRARLDAFLEGLQTVIDRHVLR	2179
QY	14	-----	13
DB	2180	SVHWGLPQAVQVVHQAQSPHITLTLRPDEDALSQLDRSLDPLRLDLRQAPLLLAYI	2239
QY	14	-----	13
DB	2240	ARPDSERWLLALIDHHMISDHVTLELIEIRLMRGSAELLPPQPYREFVAQTLASP	2299
QY	14	-----	13
DB	2300	SSAHEYFTORLADVDSPTAPFELLEVOGDNVDESELSLSDLCARIRTQARCGMSP	2359
QY	14	-----	13
DB	2360	AVLFHVAQVLA RCTGRDDVWFGTAVTGRLOGLTGAERAMGPMNTLPVRVQLATQSVQ	2419
QY	14	-----	13
DB	2420	ELVNATHRDLSELLSHBQASLAAQRCSVATGVPLFSSLNRYRHQNESOLOQPGRLLL	2479
QY	14	-----	13
DB	2480	DSABRTNYPCLSVNDYGSGLGILLIHVSQPADPQRLCAAMQCALEQLTALAHTPQKEVT	2539
QY	14	-----	13

```
Db 2540 QLDVLPAAERNLLLETFNQTRODYPTDLCIQHLFEAQVQTQPDATAVAFQAQRLSYAELN 2599
Qy 14 -----VGPD----- 17
Db 2600 ROANRLAHLHILGIGCPDPRVAICVERGEMVGLLVKAGAAVPLDPAYPAERLAYM 2659
Qy 18 -----:||||----- 17
Db 2660 IEDSQAALMTORHLOEYLPFTLTPLVLDDQDKTFTTERDDNPVVEALGVNRNLAYVIY 2719
Qy 18 -----:||||----- 17
Db 2720 SGSTGNPKGVMIHRGLVNYSDAARLFDLSPTDVTLOQNTLNFDSVEEIPALLAGAT 2779
Qy 18 -----:||||----- 17
Db 2780 LTPSREIFGSEGTENHGINPTVLHTAAHHTLVAEWHKQPOVABORLOHVRLINVTGDA 2839
Qy 18 -----:||||----- 17
Db 2840 LSAQKLLWDEVPRPAHTRLINTYGPTEATVCTAAYVSHDAAAGSEGSNATIGKPMANT 2899
Qy 18 ---VFOAQOE-----D 25
Db 2900 RIYLLDAHQOPVPGVAGEIYIGDGVARGYLNLEEVNAERFLADPFSESPPARMYKTD 2959
Qy 26 TERYV-----:||||----- 30
Db 2960 LARYNADGRIEYLRNDFQVKVGRFRIELGEIARLGNCTGVKEAVVIAREDNPGDKRLV 3019
Qy 31 -----LT-----:||||----- 32
Db 3020 AYVVAQOSOLTAADRLAELAPOLAEYMLPSAFVLLDELPLTPNKKLRKALPDADAL 3079
Qy 33 -----:||||-----NIN 35
Db 3080 ISRGYAPOGETETQIVAIWQDLGIEQVGRHDPFELGCHSLLASVLIERLRKQGLNIN 3139
Qy 36 I-----:||||----- 36
Db 3140 VKTFTAPTREMALAIQDKHVLFOVPANRIPAHCTQTPDMLPLVELSVAQIELITSA 3199
Qy 37 -----:||||----- 36
Db 3200 VAGGAANIQDIYPLAPLODGIILFHYLLNRERDAYLMRSMIEFDSRARLDAFLEGLQTV 3259
Qy 37 -----:||||----- 36
Db 3260 RHDILRSSVHNTGLPQAVQVVRQAQLPVHTLTLPTEEDALSQDLRLSDPGRLLDLRQA 3319
Qy 37 -----:||||----- 36
Db 3320 PLLAYIARDPNSERWLLALIDHMMISDHTVELILEIRLLMRGQADLLPPQVPRDFV 3379
Qy 37 -----:||||----- 36
Db 3380 AOTLASPSAHEAYFTRRLADVSPPTAPPELLEVQDGNDEEAKLALNSDLCIRTOA 3439
Qy 37 -----GAE-----:||||----- 39
Db 3440 RERGMSPAVLPHVAAQVMACTGRDDVVFGTVVTORLOGTAGAERAMGMFMTLPVRVQ 3499
Qy 40 -----:||||----- 39
Db 3500 LTTQGAQLVMATHRDLSELLSHEQASLALAQRCSSVATGVPLFSSLLNRYHQEDNRQ 3559
Qy 40 -----:||||----- 39
Db 3560 WPGMRLLDGTERTNYPLCLSVNDYGSELDIIHSMQPANPQRLCAMMOCALEQTLDAHAH 3619
Qy 40 -----:||||----- 39
```

```
Db 3620 TPQMAVTQOLDVLPAAERNLLLETFNQTRODYPTDLCIQHLFEAQVQTQPDATAVAVOGOR 3679
Qy 40 -----:||||----- 39
Db 3680 LSYADLNROANRLAHLHILSIGIVPDDRVAICVERGEMMIGLLVLKAGAAVPLDPAYP 3739
Qy 40 -----LLRD-----PSLGAO- 49
Db 3740 AERLAYMITDSQPAALLTLPGLQDRLPALSPMLPLVLLDDEQYOGLAECDDNPVWPLGV 3799
Qy 50 -----:||||----- 49
Db 3800 LAYVIYTSSTGNPKGVMIHRGLVNYSDAARLFDLSQSDVTVLQONTLNFDSVEEIPF 3859
Qy 50 -----:||||----- 49
Db 3860 ALLAGATLAPSREIFGSEGTETHIGIOTVHLHTTAHHTLVAEWHNQPOAAEORLOHVRL 3919
Qy 50 -----:||||----- 49
Db 3920 INVTGDALSQKLLWDEVPRPAHTRLINTYGPTEATVCTAAYVSYDAAAGSEGSNATI 3979
Qy 50 -----:||||----- 49
Db 3980 GKPMANTRIYLLDAHQOPVPGVAGEIYIGDGVARGYLNLEEVNAERFLADPFSES 4039
Qy 50 -----FRVHL-----VKHVL----- 60
Db 4040 RMYKTGLARYNADGRIEYLRNDFQVKVGRFRIELGEIARLGNCKGVKEAVVIAREDN 4099
Qy 61 -----:||||----- 60
Db 4100 PGEKRFVAYVVAQPOTQITAEALRAELAPOLAEYMLPSAFVLLDELPLTPNKKRDKALP 4159
Qy 61 -----TEPEGAPNIT-ANLTSSLLSV----- 80
Db 4160 APADDAFASREHVEPOGATEIALAQIMQSLDLERVGRHDPFELGCHSLLAMRLISOAR 4219
Qy 81 -----:||||----- 80
Db 4220 HOLGVELGALAFHAPEVSAALAVIAOAGRSKFPDIPVVARDOAWPLSFGOORLWFLAQM 4279
Qy 81 -----:||||----- 80
Db 4280 EGASAAHYHYPAGLSLHGNLNLKALQALERIVARHEGLRTTFMQDDGQPVQIRISPADTG 4339
Qy 81 -----:||||----- 80
Db 4340 FNLQMHDLQGLADAEKQLALASEESLQSPDLOQGLIRGLIRMAEDHHVLLLTLLHHIV 4399
Qy 81 -CGHS-----:||||----- 84
Db 4400 SDGMSVDVLTRELSALYAASFQSQDDDDPLAPLEQYLDYAVWQRRWLSGDVLOQOSNFWQ 4459
Qy 85 -----:||||----- 84
Db 4460 TLADAPALLMLPTDEARPALQDYAGAAALPVVFDKDLTRGLKALSORRGSTLFMTVMAAWA 4519
Qy 85 -----:||||----- 84
Db 4520 GLLGLRAGQDDWVIGTVPANRTRSEVGLVGLFVNTLAIRVDSLDPKPTAETLLARVKOOT 4579
Qy 85 -----QTINP-----:||||----- 89
Db 4580 LDAQAHQDLPFEQVVEVINPLRSLSHSPIFOAMLSWENNEASDLTGLDNTLKSIELA 4639
Qy 90 -----:||||----- 89
Db 4640 AQFDLTLDMAEVDQOLVCTLEYATATLFDSTMRRYLVGYFORVLEAMVADDQVLEHVALL 4699
Qy 90 -----:||||----- 89
Db 4700 GADEREHLLAGLINATEAPPQDRTIHOLFEBRVOAQPDIAVAFGAORLSYAEINRQANR 4759
```

```
QY 90 -----EDDT 93
Db 47%0 VAHHLISLGKPPDRVAICVERGEMLIGVLKAGAAVPLDPAYPAERLAYMIEDST 4819
QY 94 -----
Db 4820 PSALLAQORDVQAHPLTLDLPLVLDEQRTTLSERDDNPVETLGAHNLAYVIYTSGSTG 4879
QY 94 -----
Db 4880 VPKGMVMEHRLFAVSAANQOLVALHAPLNLHQLQAGFFDVFSADLIRSLAFGGLVLCPR 4939
QY 94 ----DP-----GHADLVL-----
Db 4940 ETLMDPPALYRLISEESIGFADFVLAVLNALLGHWVEETHDLSFMRTVVCGSDIWTASHA 4999
QY 103-----
Db 5000 ROLRKLCDHVQVQAVGVTEASIDSTCFEATSQVDGVLPIGRALANTRIYLLDELQ 5059
QY 103-----
Db 5060 AVPPGVAGELYIGGAGIARGYLNPOLTAERFMDNPFVAGERLYRSCDMARYRADGNIEF 5119
QY 103 ----DP-----YITR-----
Db 5120 LGRNDSQAKLRGLRLELGEIEARLAEVAGVRESLVIREDSGGTPKLIAYFVEYATRDES 5179
QY 107-----
Db 5180 GPALTPRALRQQLNLPEYMI PAAFVRMAALPLSANGKLDRRALPEPDADAFDQHDFA 5239
QY 107-----
Db 5240 ADGPLETAIAIADVLGVAQVGRHDDFFALGCHSLLVMRLAQRVQROLNLEVSFVFA 5299
QY 107-----
Db 5300 APVLROPAERLGNQDNARVAIKVQRSAGALPLSYAQRWLFLAQLEGGSAAYHIPAGLR 5359
QY 107-----
Db 5360 LRGNLDOASLORALDRIVARHEALRTTFVQEQGPAEORISAAETGFRLOLOVLAGOTDA 5419
QY 107-----
Db 5420 EDKLLATAAQEASEHFDLVNGPLVRGRLVRMANDNHVLLVTMHHILSDGWSGTGLNREL 5479
QY 107-----
Db 5480 ALYAPFRQAGDPLPALPVQVVOYALWQDWLSGDVLQOQRQYVQQALAGAPALLTLPTD 5539
QY 107-----
Db 5540 RPRPAQDYSQTLLELVDLTOLTRGLKALSQRHGSLSLFTVMGAWAALLGRLSGQDDVVI 5599
QY 107-----
Db 5600 GTPVANRMSAEVENLIGFFVNTLAIRVDLSGTPSVEALVRQVKORTLAAQANODLPFEQV 5659
QY 107 -----FDLE-----
Db 5660 VEVVOPQRSLSHSPIFOAMLSWQNNEDTALVLGDLTLQGVAVAGDTAKFDLALDIGEVDG 5719
QY 111-----
Db 5720 QLIGTLEYATALFDESTMRIRYGYFLRLLEAWAADDQVLEQVPLDPTAEREYLLKDINA 5779
QY 111-----
Db 5780 TERTYVQGMHRLFEAHAEAAPOAIAVRQGEQTLTYAELDSRANALAQHLRKHGVGPGT 5839
```

```
QY 111 -----
Db 5840 RVAILLDRSVELLASMLATLKGAAYLALDRLAPEERLFMLEDSEAIMLSRDLTAPD 5899
QY 111 -----
Db 5900 MTPRLDLDLLELSALNOEPVPVATEVAGETPACIIYTSGSTGVPKGVIVTHNGIVRLVOD 5959
QY 111 -----
Db 5960 NGYDFSAEDRVAFSSNPAPDASTPEIMGALLNGCQSVIIIEPQVLLPEVAFALLKRHGV 6019
QY 111 -----
Db 6020 TAMISSTALFNLYAGLIPALAGLRMIMCGGERADPASPFRVREHSAQVRLFNQYGPTEG 6079
QY 111 -----
Db 6080 TTCATSYEIFDVLPTLSLPIGKPNANVRYVLDARREPVMGVVGDIIYIGTGVALGYL 6139
QY 111 -----LPGNRQ-----VRG-----
Db 6140 NRPELTAERFSEDPFSQAGARLYRTGDLARWLPGCNLEVLARNQGVKVRGPRVELGEI 6199
QY 121 -----
Db 6200 ESVLHLYDGRNVSVVVAHEASPGDTRLVAYTYVHAGVAAAPDPEDLRAQLSADLAEMYVPS 6259
QY 121 -----
Db 6260 VFVRDLALPTLNGKVDQKALPVPDLNAMFERTYVAPEGATEQALAEI FOELLGLSERVGR 6319
QY 121 -----
Db 6320 HDGFFELGCHSLLAAQLVSRVQQLNGDMALROLFNHPTVAELAKVVDGLQTVDSDSIEP 6379
QY 121 -----
Db 6380 IERNAPLALFSOORLWFLDRDPCGASSAYHMPMSLLLRGELDRALKALDRVARHES 6439
QY 121 -----
Db 6440 LRTTFELHGEQPVQVITAAADSGFALAEDDLRSQPYEQASLNASRIADSEANAAPFDLROGP 6499
QY 121 -----VTQ-----
Db 6500 LIRGRLLRLADDEHMLLIITQHIIISDGSVGVLINEFTALYQAFTEORPDLPLPALSIOYA 6559
QY 124 -----LGGACSP-----
Db 6560 DYAAWQRRTFTGERLAEQADLWREHLGGA--PTLLSLPTDRPRPVVQSVRGGAVPVTIDA 6617
QY 132 -----WSCLIT-----
Db 6618 ALHQRLERFCQAHNVTLFMGLLSAWSVLMTLRLGNERDQVIGVPSANRGRTERETENLIGFFV 6677
QY 138 -----
Db 6678 NALALRVDLTONPSVAQLLEQVROTTLAAHEHODIPEQVIEALQPPRSMHSHPLCQVAL 6737
QY 138 -----
Db 6738 SLDNTSIGGELKLPGLSLHPVLOAHETAOFDMLTLASENGALNGVIEASDLFDRSTVE 6797
QY 138 -----
Db 6798 RFAQHFTLLEAMVEDVAQPVGLPLLSPAQRLASPALLQPKAVFASGLMVHORPEOFAA 6857
QY 138 -----
Db 6858 AHPONTALVFRGHEVSQALNRRANRLAHELLAAQGVPRPDRVAILAERGTCQMICAVLAVL 6917
QY 138 -----
```


A:Accession: T15789
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7829 <BEN>
A:Cross-references: EMBL:U41541; NID:g1109867; PID:g1109869; PIDN:AAA83181.1; CESP:C41A3
C:Genetics:
A:Gene: CESP:C41A3.1
A:Introns: 29/2; 69/3; 141/2; 192/2; 271/3; 321/2; 1000/1; 1098/1; 1127/3; 1168/2; 1282/2; 7514/3; 7596/3; 7654/3; 7716/3; 7732/3; 7769/2; 7797/3
C:Superfamily: acyl carrier protein homology
C:Keywords: carrier protein; phosphopentetheine; phosphoprotein
F:6547-6616/Domain: acyl carrier protein homology <ACP>
F:2832,5271,6580/Binding site: phosphopentetheine (Ser) (covalent) #status predicted

Query Match 30.1%; Score 233; DB 2; Length 7829;
Best Local Similarity 1.3%; Pred. No. 65;
Matches 90; Conservative 25; Mismatches 33; Indels 6699; Gaps 27;

QY	1	AAGGI-----	5
Db	151	AAVGIETACSSSLVAFHARQAIOSETKLALVCGANHVGSRFSHLYNHMVSPNGRLA	210
QY	6	-----	5
Db	211	AFDRSANGFVRAESFAVAVLCSKQFAEENLLIHCECVGSFNSDGTPTAPNPISQY	270
QY	6	-----	5
Db	271	EVOLEAKNIDKDSVOLVTCGTGTLGDOVELTAINRSFKSDIRVMSPKSMGHGEGAA	330
QY	6	-----LHLEL-----	10
Db	331	GLIGVLQSLYSMQHGIIPNQLHLELPSEDLGEDKSMGFVNBEMELNRVAISSYFGGTNA	390
QY	11	-----	10
Db	391	CAIIEKPKSLVQKESYAESNVFLPSAKSHESLKLQIEEYTFQMAQSDAMEDILTYVN	450
QY	11	-----	10
Db	451	ERKTYDFAAVFGKDNIEIARKLQDGDYSLTNIQESTFEVEFGEGNEKLWLLRMLYEKN	510
QY	11	-----	10
Db	511	ETFHSTVDKYCKLAETCGFPEARTALFPFKLTLTPLTYNVSRLLISSMATFELLVQYNTL	570
QY	11	-----	10
Db	571	PNKLRGKLGQIFCLAVAKVITFESAVOLIKGVVAEANLTDILGDIELKSKKIPIQHL	630
QY	11	-----	10
Db	631	KSTKKKILPIHISGELKETAKPNLMTFVNGEILELDPVRVKOKLICOLFACGFDPAVK	690
QY	11	-----	10
Db	691	FRGRIVKPTYSFLKQFQWPEVQMTATNIVDQTNSSLSDAEISTVRTIVKQFLDIEE	750
QY	11	-----	10
Db	751	DDINLLETGAVDSLTSIEMVEAFGTAVNQTMPFDLLLEAYPTILNIVDFLKLTVTPVTK	810
QY	11	-----	10
Db	811	ATTSIHKTSELSTSDINVIACDYQFAGVEGEBELWDLTTLTSLRTTGKISDIRKKQCEGD	870
QY	11	-----	10
Db	871	AGLEVGLLKQDISMPDNSFFATAKOEAEFLDPQHRLLNAAYNALKSGLTSIPDADFL	930
QY	11	-----	10
Db	931	AISAHSEYRALAEKHINELDERLWGTVHSMVAGRLAVLMGIRGAMIVDTTCSSVATAL	990

QY	11	-----	10
Db	991	EMAVKSIREGRKPAIVATSOLOSCKWLSYSLKTLDDHSTNSFVDCSGFCRSDGVGVII	1050
QY	11	-----	10
Db	1051	LKTAEGDSAVIKISSAKSHHCGAVMTPVSSISQLEENGSGSYVEGHGTATSAGDSAE	1110
QY	11	-----	10
Db	1111	SMAYQKLGSELIMSSVKAQFGHCEVAGSLQLMKVSSIGHGIIPISVHNILPSEHIRNN	1170
QY	11	-----	10
Db	1171	ENIRLPFAEEKQIDRSAIVSFGITGCTKTVVTVTERVSQLNVNIONCYLLPVSAKTKDGL	1230
QY	11	-----	10
Db	1231	KKACLSLIEIMDNSCESLYDISTTLOKQKTNPKWRTAVVGSSHADVVLKQKFLTSEHT	1290
QY	11	-----LVAV-----	14
Db	1291	SLTNWHISTSHHSIGCSITFFHNIPEFEDHYSMFCHRLRLEFEPHSNTNSIYHLLAVVYALIR	1350
QY	15	-----	14
Db	1351	VILKHLTNSFAVGFSGLIVLAAVDAAPSHYLNLDLLHAFANDDDVRMMKRIARDVTISIE	1410
QY	15	-----	14
Db	1411	NVKLLNNGEPTTARQAVEATIDOKVKNRLEPETTLILSPRAYEFASOLETTIDQYVKL	1470
QY	15	-----	14
Db	1471	IGKEFTGGQVDFAGIFGTPIKLDLPEYPNRSFWLPIDPSVPSKEGKELIPKSYE	1530
QY	15	-----	14
Db	1531	FLLSQKQKHQVQNHVDSKIVLPGATSIRLVHOLNGKPTVELSNIDFLNKITPSEAPSVV	1590
QY	15	-----	14
Db	1591	KIBEQDGLKLVFGETAISFKLTELONFNPIPERLNABEVHTDNIYERFANSHLYTRN	1650
QY	15	-----	14
Db	1651	EFQVDSLKYTMGKEVRFVSMKDLIDLIDGTLQAIVCVCFPENTNDNSPFVPFTIDOLS	1710
QY	15	-----	14
Db	1711	ILNGDISOKOLHAVLYKVDSSGNFNGDATVYDALGNILHISNVTFKRLNGOSAPSLTSK	1770
QY	15	-----	14
Db	1771	TVDSKITKKVENEDOKRASKNNLHVWFENFGWTDIDNTTGFDFDLGLTSIQAVKLRAI	1830
QY	15	-----	14
Db	1831	KSNYPNASSTCVDPYPSIDLLSGYLSTLNDPOVTETSTGEDDIOKDLTEDHKPTRAENP	1890
QY	15	-----	14
Db	1891	IGVMAAACRLPGGVSSPSSELWELLKIGKNASSRIPATRVPTNTLISGSKYGNPVEGNF	1950
QY	15	-----	14
Db	1951	ITQDVTQDPDPFFKISKSEBELIDPQORLLLECVQCLENSGVETSNVGVFVGLMEKEY	2010
QY	15	-----GP-----	16
Db	2011	QDMMESSILAMLGSMMAVIAGRVNYIFGCYGPSVTIDTACSSSLVALEMAINALLDNRC	2070

QY 17 ----- 16
Db 2071 SKVIVAGVNLILNEKGOGLRTNGKMLSOHGMSLSPDGRSGYGRDGCWVLMLELAKPNF 2130
QY 17 -----DVFOAH----- 22
Db 2131 HYMTIQSVNVNHGGRSVSLTAPNGVAHKMLLTSVINOSPSLAIDYWEAHGTGTPGLDPI 2190
QY 23 ----- 22
Db 2191 EFNTLSSTLQNIIGSVKASLGHGEASAGTCGLLKLFLMLTYQVPTLIHFHVLNKDINA 2250
QY 23 ----- 22
Db 2251 GSIRLPIIGEDSELVSAGISSFQVSGTNAAIAFNDNNKLEPYPIHKKYILPISAKNOI 2310
QY 23 ----- 22
Db 2311 SLDNLEKQILSVIPLTDVPICNIASALANRSHFTIRNALIVSNSGIVNSKMEGKPHVA 2370
QY 23 ----- 22
Db 2371 KKDRVHVKCLDSDLASLLQYDVINETYVASLKNPQSFAMKFAIKELTSLSEYIEIVA 2430
QY 23 ----- 22
Db 2431 SDGELLAVLLANGSLKWFNFKTWIELPIGSLLTEFADHDLNSTSSIKSYOTQPESH 2490
QY 23 ----- 22
Db 2491 NLDSPMELMKLMKIYITGYDWDATVYSPVEQFIALPNYQFNKOTLWFEERLEIVDHYL 2550
QY 23 ----- 22
Db 2551 IGTIDEESEDTLILKNOISELRHPQFFKGKPLDVGTMSEIAIEALKIRNEIPFSIQNLKT 2610
QY 23 -----QED----- 25
Db 2611 ELITLTPAWLETNVTNRNEDDEGFNVSAIYDQGRUFLSNASSVEIQNIEVPAVEVQIPDK 2670
QY 26 ----- 25
Db 2671 VVYLKECPNAVIRRHNMVYVDSRAEQSPFRRTANIVLNEIGFAPTPSDMFIELIGVLPS 2730
QY 26 ----- 25
Db 2731 VHYMYQVDDGALWQFMISQDKRVLNIVYLKDAKGLEIPTIRMHKKSTLLSSQEAIVA 2790
QY 26 ----- 25
Db 2791 AKTLQMAVRHKVCLAVGDVIESGLDIDESQLSTGFSELGIDSLATVDLLNRLNKYFPEI 2850
QY 26 ----- 25
Db 2851 ELTTSOLFNDPSIIDLSIMIEOLLNEKGITEPSEBPTKTSRGRKLSIPAVRAQVLAQI 2910
QY 26 ----- 25
Db 2911 EFVENYNSKQKEVQAEAPSSSECSNHLSESDATVDRTEIRRKVSLAVFDLATETLSAED 2970
QY 26 ----- 25
Db 2971 LOSKGFTELGMDLSLIVDFVNRNDKYFPDDEITASDIFDYPTVDELSDHIVRKKSSVP 3030
QY 26 ----- 25
Db 3031 PAASEIMKETMNGISTSVDAETHKLENLSQSPMLLENQNSINPTLXMIWSNOTIKLVKPS 3090
QY 26 ----- 25
Db 3091 DGNFLFELNANGQOEKEIQKHFTGPNNIIDLKGFHEGSTETLYMNLNVLKVSISKLEIQ 3150
QY 26 ----- 25

Db 3151 CRFGVSOEFGLGNSISRAFMKTVAAEKNPLISFANYQNVQVSDSDSPITGNWLTGG 3210
QY 26 ----- 25
Db 3211 LSGIGLEIGKFIANGAENVILISRROPTAKALRDILSTELTHIGLARKTIVLKIKNIS 3270
QY 26 -----TERYV----- 30
Db 3271 AKLIVFOSFKLSPSTKKYISDLFSKKVTPFPYNLLQSKCFKHFHWFLEWLEVYLKQWTV 3330
QY 31 -----LTNLNIG-----AELLRD----- 43
Db 3331 HTAADINDKEKLIRELTKLNVGITGIIHSAGVLKDSKIERONKESPNQVTPKANGPHV 3390
QY 44 ----- 43
Db 3391 LEETEKHFNYKIENFIMMSSFTAACGNEGQLVGSNAYLEYOVQRRRRROCKSGCAIOMG 3450
QY 44 ----- 43
Db 3451 NWIDTGMATDENVRKFLUANLGLQHNKDALKYLKACILTKPELIMVANIDWNVILKNRK 3510
QY 44 ----- 43
Db 3511 DLPKDLINTGILPEDFTGKNESEFPPLSNQDPKESVMNFSVEDEEVELELIKEKVSSIL 3570
QY 44 ----- 43
Db 3571 MCSPTLKNKNKINMDGLDLSKLIVELFNFINSTFKISVNLSDAYNHPTLEKLAHAHFEOM 3630
QY 44 ----- 43
Db 3631 TIVDHPVNSVKSEBIFKSTDFCPIFGINIFPDNKNDFDPAKSTAVKLENQOLPTAGKY 3690
QY 44 -----PSLGAQ----- 49
Db 3691 AVSVVGSIRDVSVKIEAAPQIKLCQESSKCVMLMTGGSGOYPMHGRQLVENYEIPTS 3750
QY 50 ----- 49
Db 3751 TLQSLCKCDEYLQGVSLWEILFNTDHYKLLQLTQHMQIPMFCGYATAQLWLSLGIPT 3810
QY 50 -----F 50
Db 3811 DYLLGHVSELVAGVLGAGIMSIEDGLRLIVERGKAMENIAGLALLAVQREIADVLKRP 3870
QY 51 RV-----HLVKM-- 57
Db 3871 KVSATINSKQVVFAGTKSVLDAALAFVKQGGQKQATVYNQYPPHSHNLQETHLVLSURQ 3930
QY 58 ----- 57
Db 3931 CLADIKFSAGRTPLVSNVTGOIINTFSEAYIVKHVTSAVKFDVCVETLQAKGVTVMIDAG 3990
QY 58 ----- 57
Db 3991 SAAVLATFVKRIIOPTELSKHRIVOTCKEKESVDNLVQACLEQSLPISWTTLYCGG 4050
QY 58 ----- 57
Db 4051 RNADERLVEFPMTNNDIIKNDEFELLEGHQNGKIVVAGAYQLFKIDQLVKLKAAGMELM 4110
QY 58 ----- 57
Db 4111 LKNVFLKPMYIEDNREYQIQWNSDMTIELIVNSVIVGSELEPQNSVLKLETISENEKP 4170
QY 58 ----- 57
Db 4171 FEVHDFVETLFRNGLOVDSGFRRIESARRSDKRCFSQIKSPPFAWPLIDSAMHSITASVW 4230
QY 58 ----- 57

Db 4231 PRPDYFLPVAMGSVTMKTNSFTPLNLHAQTVITSETDKFIQNVALLAGDTPICEVR 4290
Qy 58 ----VILTEP----- 63
Db 4291 NMTIVLUKTEPVHTRIPNSIETVETPPKSEIEIVGFDISLPYNOISENENWHLKNTV 4350
Qy 64 ---- 63
Db 4351 KQLHNRSLKQDHARVALLDSADARYWDEYFIRPSEAKFIDPOORLLCSVAKLLDSSL 4410
Qy 64 ---- 63
Db 4411 ITSLSNTGVFIGSANEFSHIVAYGYKOPRAEWSGGTNSALAGRIAHWLKLGPPVT 4470
Qy 64 ---- 63
Db 4471 LDTACSSSFYALSACDALRTGQCEYAIIVGTNVLMHEMTDVLQNAKMTVDDFCKAFDV 4530
Qy 64 -----EGAPNITA----- 71
Db 4531 DANCYKSEAVCSMLLTSPNIDSVAITNYATGHNGTSSSLFTPNGLSQLEVMQATNP 4590
Qy 72 ----NLTSSLLS 79
Db 4591 LEKILEIQTHCTGTGLGDPHIEINAIKLVSSACKIGSVKSNIGHTGGSLVSLCSLLMS 4650
Qy 80 ---- 79
Db 4651 FRSKYVAQLHLKPTNSIKTNMKICRFIGEDADENNILINFGFTGSCNCSVVLPKPN 4710
Qy 80 ---- 79
Db 4711 ISEHFVSSEVPYILLSSHSKSLQKYVOVLCEFISNSAKSLDHIMMSLFQKKIHVRQF 4770
Qy 80 ---- 79
Db 4771 IIFNFKRIAVTSLDGEVFRDERLEKLKHPCSVFLKEGVVHFDKSKFORVDLPSIVFN 4830
Qy 80 ----VC----- 81
Db 4831 NTLHWALDSYRDEIDRHESQMSPKNIFYEKVLLETMPQONPKSVKVCIGRDLIPKIE 4890
Qy 82 ---- 81
Db 4891 IDEVSSFANGIIIVPHIPISNIFEYLKAKVWSLIRNQNVFIICCENGTSHTWTGTL 4950
Qy 82 ---- 81
Db 4951 RSLASERMPYKVSIDKVDALKLEFNHEYMFEAIFYSKRGYVERLRKVNPTKFAPOY 5010
Qy 82 ---- 81
Db 5011 EKCLISGTTGIGSAIINELPKSSVITRKNIASBDKTFELSSDITRLDISHKFNVVFH 5070
Qy 82 ---- 81
Db 5071 LAGIVNNSLHENVKRDSDLENVSIKQAKNLMKCCDTSHTFVSSSIANVLGSGOSNY 5130
Qy 82 -----GW----- 83
Db 5131 AFSNGLVTSFLETSTKSTIIHWGPWKDVGMQAQPERREIVKQIESNGWKLLPNQDAISV 5190
Qy 84 ---- 83
Db 5191 FYOTFMETHEOIIVFDGGFTIVARQPHLOKLLSEVVEKTKVKEEIKKKSUNFEEIFEEI 5250
Qy 84 ---- 83
Db 5251 VGITDISSKLNIPEMDLIGDLSLCHMENLYSLNKNFDLELTVSEMFENATYQKLOTYYVETL 5310
Qy 84 -----SQTINPEDDT----- 93
Db 5311 RKAKHSESLSHDRVSSQVSNKEDDTRAVAVIGWSAEFGSSNIHEYENLMDGICSTGNKK 5370

Qy 94 -----DP----- 95
Db 5371 YLLKNPFGDNKFPNLTDEDDARVLDPOVRKFIOHAYLALENSQVVKQKHELRCGVFAGAE 5430
Qy 96 ---- 95
Db 5431 PSDYGRADDDHDAWRKLFVNMNMSYLASYASYCLDLKGEAVSVYSACSTALVAVANAVKS 5490
Qy 96 ---- 95
Db 5491 IQSGMDYALVGAASIAEVSGALSGDFDQKKTWFSKSGVCRPFDDKXSEGIVRSGVGCFFV 5550
Qy 96 ---- 95
Db 5551 LKRYSOALLDNDNVHVIKDFAINNDGHSRASFMAPNAGOLKCMTDVLFARFTNKEXER 5610
Qy 96 -----GHA----- 98
Db 5611 SFVECHATGTTGLDTIEMNSLRTAYSFKNKLAIKCKCANIGHAYASGLAALVKCAKMLQ 5670
Qy 99 ---- 98
Db 5671 TGIIPQVNFSEPRDGMGOFTVNGKKSTISONSLISIDSGIGGTNVHVIFFPARSQE 5730
Qy 99 -----DLVLY----- 103
Db 5731 VWKISSENILYDMIPISAKTEYSLOHTSEAIKSYLOTDTNKIAQCSSTFIHSRVPMSDR 5790
Qy 104 ---- 103
Db 5791 TYLSVNNNNELLKIRTNKKTWFKNGSPKIALPFAQOIOFTNLPNEYLNKSAYRREVEY 5850
Qy 104 ---- 103
Db 5851 LCELASSFGIPSLLEGILYPTKNFDHLIHATQFAQIAIFVOCMAIFKAIKVNVPNTCLIGH 5910
Qy 104 ---- 103
Db 5911 SVGEYAAAVISGALKTEEALKLLIKRSELLIGKTEKARMLMVWNYEKOLPSHVHVSALIDA 5970
Qy 104 ---- 103
Db 5971 NTKCVVPVETIONLEKYFINNHIKYRNIEHTKHGPHSKMPCISKPEFFCESFATKVP 6030
Qy 104 ---- 103
Db 6031 IPMISSITGSEIKIFDSKYCTMHLTNPNVLELVVDHIMKLDIDIIVEVGPTGVLSNLLAK 6090
Qy 104 ---- 103
Db 6091 RNSKI VVVPTCGTKKHPKISLGECIGQLHNGVDIRKLTPLKAVDGOVPGYCFCDERQFGE 6150
Qy 104 ---- 103
Db 6151 KNQKISQKLNRPYSYKESKMPVSSELEETKPVRTVCDGSLNGISEPVLVLLRKSDIH 6210
Qy 104 -----ITRDL----- 109
Db 6211 KNYFDVELFLKSPRCSTVVFGMDNSPSVHLSGLIRCYQLVSRIDLKYVENPENIAMS 6270
Qy 110 -----ELP----- 112
Db 6271 IVIQKVLTLANGLYFRIDSTCLCKHGFRTEIPSRLHPVSRGHALVPGANGFIGSIVFRLL 6330
Qy 113 ---- 112
Db 6331 QEMGMNVIPISRASIPSCDITNIKDVONVFKSLGFKKFSVIVNCVGVETSAKMNKTSLEQ 6390
Qy 113 ---- 112
Db 6391 EIVLSPKTFGSVNILKLEEFSEIVDKLVNFSSLSVVPLLGNFDYASANCFFVEALTQOG 6450

```
Oy 113 -----DGNROVRGV----- 122
Db 6451 SKYIKOFLTLSPLEGRMYESSSTKHNLSQILMFQEBITNVKLTAPESNMGIVFV 6510
Oy 123 ----- 122
Db 6511 SSVNPADIATKSTAFHRTDENSEVKAVESLKLPKSTCEPFAIEIWKETLGISILNDAN 6570
Oy 123 -----QLGG----- 126
Db 6571 PNFFSLGGDSLALQVVKVQKKTDRIVDVNDLFDNPTLQEFKTFVKNLTTEKFAGNTND 6630
Oy 127 ----- 126
Db 6631 KISYDAIPLTNSQTMFMLRQIQTTSKYNLIPIKITISYETKFWBFLKYSLSLHSLIAYOPS 6690
Oy 127 -----ACSP----- 131
Db 6691 YRTVFKSGNSPYQYICSLTESPHDFDKRCNNNAISHEPNHLFEICKSTPLRVRVAECD 6750
Oy 132 -----WCLITEDT----- 140
Db 6751 NSRIHIVNQHHILTDGWSMTVLSDTVSSLYAAYRGTSFPFSKTKQTISQVAMGKSSGD 6810
Oy 141 ----- 140
Db 6811 IKALEYQNTYHTIIPYDSETGNTSPSYVRISKLIPIKIPSKLIPKIVOKLVGLSKLYNTTMYNLAL 6870
Oy 141 ----- 140
Db 6871 SVPCDAVRSFTGQADILLAYAGRNADNSSELIGYFMNNAIPKTSLPFEILRLEILNIV 6930
Oy 141 -----G 141
Db 6931 LNSLEKSRSFATIPFYQMVQNRKLNELISLFFNFRQKLDYPTVSMFGAKCEIEHLNLNA 6990
Oy 142 FDLGVTI 148
Db 6991 FDFSFTI 6997

RESULT 7
T21371
hypothetical protein F25H8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21371; T24896
R:Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19413
A:Accession: T21371
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <W1>
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: Clone F25H8
R:Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19949
A:Accession: T24896
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <W12>
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: Clone T13H10
C:Genetics:
A:Gene: CESP:F25H8.3
A:Map position: 4
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 81
```

Query Match 29.6%; Score 229; DB 2; Length 2165;

Best Local Similarity 3.7%; Pred. No. 1,2;

Matches 61; Conservative 20; Mismatches 65; Indels 1522; Gaps 8;

```
Oy 3 GGILHL----- 8
Db 5 GGSFHLQPVVAALILLVCLVYALQSGGTISEFSSDVLFSRAKYSGVPVHHSRWRODA 64
Oy 9 ----- 8
Db 65 GIHVDSHHIVRRDSYGRGKRDVTSTDRRRRLQGVARDCHACHLRLRSSDAVYIVHLH 124
Oy 9 ----- 8
Db 125 RWNQIPDSHNKSVPHFSNSNFPAMVLYLDSEEVGCGMSRTDPDCIYRAHVKGVQHSHV 184
Oy 9 ----- 8
Db 185 NLCDSEDLGYMLALPSGIHTVEPIISNGTEHGDGASRRHQLVRKPDPMHFKSFDHLNS 244
Oy 9 -----ELLVAVGPDPVQFAHOEDTERYV 30
Db 245 TSVNETETTVAWQDQWEDVIERKASRRANSWDHYVEVLVADTKMYEYHGRSLEDYV 304
Oy 31 LTNNIGAEELRDPSLGAQFRVHLVKMVLITEPEGAPNITANLTSSLVSGWQSTINPE 90
Db 305 LTFSTVASIYRHQSIRASINVVVVVKLVLKTNAGPRIITQNAOQTLQDFCRMQQYNDP 364
Oy 91 DDTDPHADLVLYITRFDLELPDGNROVRGVTOLG----- 125
Db 365 DSSVQHHQVAILLTRKDCIRSOQKCDTLGLAELGTWCDMOKSCAIIEDNGLSAAFTIAH 424
Oy 126 ----- 125
Db 425 ELGHVFSIPHDDERKCYTMPVNVKVKCFQSTKFDKTOFQNNFHIMAPTLEINTHPWSWP 484
Oy 126 ----- 125
Db 485 CSAGMLERFLENNRGOTQCLFDQPVERRYEDVVRDEPGKKYDAHQCKFVFGPASELC 544
Oy 126 ----- 125
Db 545 PYMPTCRLWCATFYGSMGCRTOHMPWADGTPCDESRSMFCHHGACVRLAPESLTIDG 604
Oy 126 ----- 125
Db 605 QMGWRSWGBCSRTC GGQVOKGLRDCDS PKPRNGKCYVGORERYRSCNTOECPWDTPQY 664
Oy 126 ----- 125
Db 665 REVOCSEFNKDIGIQGVASTNTHTWVPKYANVAPNERCKLYCRLSGSAFYLLRDKVVDG 724
Oy 126 -----GACSP----- 130
Db 725 TPCDRNGDDICVAGACMPAGCDHQLHSTLRDKCGVCGGDDSSCKVVKGTNEOGTFQYN 784
Oy 131 ----- 130
Db 785 EVMKIPAGSANIDIRQKGYNNMKEDDNYLSIRAANGFELLNGHFQVSLARQOIAFQDTVL 844
Oy 131 ----- 130
Db 845 EYSGDAIIERINGTPIRSDIYVHVLSVSGSHPPDISYEYMTAAVPAVIRPISSALYLW 904
Oy 131 -----TWS-----CL-----IT 137
Db 905 RVTDTWTECDRACRQGSQKMLCLDMSTHROSHDRNCQNVLPKQATRCMNCIDCSTWIT 964
Oy 138 ED----- 139
Db 965 EDVSSCAKCGSGOKRORVSCVMEGDRQTPASEHLCDRNSKPSDIASCYIDCSGRKWN 1024
Oy 140 ----- 139
Db 1025 GEMTSCSETCGSNGMKHRKSYCVDDSNRRVDESILCGREOKEATERECNRIPCPRVWYGH 1084
```

QY 140 ----- 139
Db 1085 SECSCDGGVGRHAQCLDAADRETHTSRCGPAQTEHCNEHACTWQFGWSDCSAKC 1144
QY 140 ----- 139
Db 1145 GDGVQYRDANCTDRHRSVLPPEHRLKMEKIITKPCHEKSCPKYKLGEMSCQSVSCEDGWS 1204
QY 140 ----- 139
Db 1205 SRRVSCVSGNGTEVDMSLCGTASDRPASHQTCNLTGTCFFWRNTDWSACSVCSCGIGHRETR 1264
QY 140 ----- 139
Db 1265 TECIYREQSDASFCDGDKMPETSQTCHELLPCTSMKPSHWSPCSVTCGSGIOTRSVSCSTR 1324
QY 140 ----- 141
Db 1325 GSECTIVDEYFCDNRTRPRLKKTCEKOTCDGPRVLQKLOADVPPIRWATGPMWTACATCG 1384
QY 142 ----- 141
Db 1385 NGTORLLKCRDHRDLPDEYCNHLDKDEVSTRNCRRLDCSYWKAWEHECPATCGTHVQQ 1444
QY 142 ----- 141
Db 1445 SRNVCVSAEDGGRTILKDVDCDVQKRTSARNCRLEPCPKGEEHIGSWIIGDWKSCAS 1504
QY 142 ----- 143
Db 1505 CGGWRRRSVSCTSSCDETRKPMFKCNBELCPPLTNNSWQISPMWCHSVSCGGVQR 1564
QY 144 ----- 143
Db 1565 RKIWCEDVLSGRKODDIECSEIKPREORDCEMPPCRSHYHNKTSASMTSLSSNSNTTS 1624
QY 144 ----- 148
Db 1625 SASASSLPILPPVSVWTSWSACSACGCRGTRKRVVCEVNPVSLNVTV 1672

RESULT 8
T36248
CDA peptide synthetase I - Streptomyces coelicolor
C/Species: Streptomyces coelicolor
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 01-Dec-2000
C/Accession: T36248
R/Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1999
A/Reference number: Z21602
A/Accession: T36248
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-7463 <SAU>
A/Cross-references: EMBL:AL035640; PIDN:CAB38518.1; GSPDB:GN000070; SCOEDB:SC563.03c
A/Experimental source: strain A3(2)
C/Genetics:
A/Genes: cdaPSI; SCOEDB:SC563.03c
C/Superfamily: acetate-CoA ligase homology; acyl carrier protein homology
C/Keywords: carrier protein; phosphopantetheine; phosphoprotein
F/516-1074/Domain: acetate-CoA ligase homology #status atypical <ACLI>
F/1090-1158/Domain: acyl carrier protein homology <ACP1>
F/1715-2184/Domain: acetate-CoA ligase homology <ACL2>
F/2200-2268/Domain: acyl carrier protein homology <ACP2>
F/2804-3249/Domain: acetate-CoA ligase homology <ACL3>
F/3265-3332/Domain: acyl carrier protein homology <ACP3>
F/4323-4746/Domain: acetate-CoA ligase homology <ACL4>
F/4762-4830/Domain: acyl carrier protein homology <ACP4>
F/5363-5786/Domain: acetate-CoA ligase homology <ACL5>
F/5802-5870/Domain: acyl carrier protein homology <ACP5>
F/6401-6868/Domain: acetate-CoA ligase homology <ACP6>
F/6884-6951/Domain: acyl carrier protein homology <ACP6>
F/1122,2232,3297,4794,5834,6916/Binding site: phosphopantetheine (Ser) (covalent) #statu

Query Match 29.6%; Score 229; DB 2; Length 7463;
Best Local Similarity 1.5%; Pred. No. 89;
Matches 86; Conservative 22; Mismatches 39; Indels 5712; Gaps 25;
QY 1 AAGG-----IL---HLEL 10
|||
Db 1576 AAGGQAGLDIAVEFSTELFDAGTVRALTERLLLLLTQVAADPGVRLGDLVLLPGEHDL 1635
QY 11 LVA----- 13
|||
Db 1636 LVAANRGTAPASPAPSPASPAPGTATILPELPEROAAHPDRALTALTFECTSL 1695
QY 14 -----VGPDV----- 18
|||
Db 1696 SYAELNARNLARLLTARGIGPDALVALALPRSAELWALLAVKVGAAAYVPLDPCYPA 1755
QY 19 ----- 18
Db 1756 DRLAHALSDSAPAAALLTDRATGRLPAHEVPRIVLDAPADGGTTGGDPADAHPTDLA 1815
QY 19 ----- 18
Db 1816 QGSRVRPLDPRDTAYVIYTSGSTGRPKGVAVPHGNVVRLEFSATAPWFGFDEHDWTLFHS 1875
QY 19 ----- 18
Db 1876 YAFDFSWMELGPLLHGGRLVVVPHVDTRDPAAPALLARERVTVLNOTPSAFHQLAAAD 1935
QY 19 -----FOAHQEDT----- 26
|||
Db 1936 RENPTELARTVVGGEALDLSRLADWYERHAEDPALVNMVYGITETTVHVSHPALDRAT 1995
QY 27 -----ERVV----- 30
|||
Db 1996 AAASSASTIGVNPDLRVVYVLDRLRPTAPGVTEGMYVAGAGLARGYLGRPALTAADRPFA 2055
QY 31 ----- 30
Db 2056 DPYAALFGERGTRMYRTGDLARRRTDGLDYLGRADQVQKIRGFRIEPEIEAVLAAHPA 2115
QY 31 ----- 30
Db 2116 VDDVAVVAREDVQDPRLVAYVYVVGATARALHDHAAGHLPOHMLPFAFVTLVDVLP 2175
QY 31 ----- 30
Db 2176 NGKLDTKALPAPAHAGVGTGRAPRPREEILCALFAEVLGVPRLTVDVDDSFGLGGHSLLA 2235
QY 31 ----- 30
Db 2236 TRLAGRIRGTLGVLSVRRLFETPTVAGLSAALDGAERSGTGTAGERPERPLPSVAQOR 2295
QY 31 ----- 30
Db 2296 LWELHQLGSPPTYNITGALRLTGDLPGALRAAFQDVVTRHESLRTVFSDEGCHGARQTV 2355
QY 31 -----LTNLNIGAELLR----- 42
|||
Db 2356 LDAAGVRFELPVADVSEDRDLARLEEAARHCFLDTTDPVRAELFRLGAREHVHLLMVHH 2415
QY 43 ----- 42
Db 2416 IAGDWSLGLIRDLAAAYTABARRAPDWPLPVQYADFALWQRAALGDAADATSPAGR 2475
QY 43 ----- 42
Db 2476 QLAHWKEALAGLPDRLELPADRLPAVASHRGGRVPLTPAPLHSGVAELARESRTSVFM 2535
QY 43 ----- 42
Db 2536 VLOQALAAALLTRMGAGEDVPLGTVPVAGRGDDADVQVFPVNTLVLRDTGTGGNPTFRALL 2595

QY	43	-----	42
Db	2596	DRVRODTLAYDHQDLPEHLVDVLSPTSLSHNPLFQVLLSLDPTQODALAALSATGLG	2655
QY	43	-----	42
Db	2656	VRLNVTGVAKLDALEIAEHRDADGAPAGLVGAAYSAIDLFDGTVTLVERFLRLD	2715
QY	43	-----	42
Db	2716	ALVADPSRRIGDGDVGLPRERERVLTENDTPRRPVQGTFFADHVARHAAERPGHLAVETA	2775
QY	43	-----	42
Db	2776	GAAPCGALTYGELNERANRLAALLARGAGPERFVAVALPRSADLVLSAALFKAAGAY	2835
QY	43	-----	42
Db	2836	LPVDPAPPAERITHLVSDAAPTLLIVTTSALAASLPDGTGTPVLLDTPETAATLAALPGHD	2895
QY	43	-----	42
Db	2896	VTDADRVPPLRPEHPAYMIYTSGTTGRPKGVVVTHTGLPGLDIFTRDCAAGSGRILQH	2955
QY	43	-DPSLGAQF- 	50
Db	2956	LSPSFDASFELAMGLLTGATLVAPPETTPGPPELAELATRAHAATHLSITTSVLGLPPD	3015
QY	51	-----	50
Db	3016	SLPDGLTLVVGABAIPELVERWSPORTMLNSYGTPTTVCSTMSGPLSGPAVPPIGSPV	3075
QY	51	-----	50
Db	3076	ANSAYVVDALRPPVPPGVELYAAAGAHARGYHRRALTAEFVANPFGEPSRLYRT	3135
QY	51	-----RVHLVKWVL- : : : : : : : : :	60
Db	3136	GDLVRMRPDQLEYLGRADTVQKIRGLRIEPTIEIAVITERPHLARAIAVIREDRPDRR	3195
QY	61	-----	60
Db	3196	LVAYVVEPGATVDTAELRAALRETLPDHMIPTAFVVLDALPLTLNGKLDKRALPAPDYS	3255
QY	61	-----TEPEG- 	65
Db	3256	ARTSGRAARDPRERLLTALFGEILGVEPAGVDDGDFDGLGDSILSIQVARARAAGLVLS	3315
QY	66	-----	65
Db	3316	VRDVFHQTPALLARSAAAAPAGDRTARDSDVPADGPAPRTPMMGWFAALGSDLAAPNOS	3375
QY	66	-----	65
Db	3376	LVLRVPAALDPDLDLTALRAVLDRSHDALRMBVADDWTIEIPPPGSVTPADCLVRFDVGL	3435
QY	66	-----	65
Db	3436	DEAAVRSVTEQARTARDRLAPADGRMLQAVMLDRGADRGDGLLLVANHLVVDGVTWRL	3495
QY	66	-----	65
Db	3496	VPDLAAAYAGETLAPVGTPIRHWALSLSDLAGORTEEELDHWHSVLGDTPTHRLDPA	3555
QY	66	-----APNITANL- : : : : : : : : :	81
Db	3556	DTHATAGEITAEADADTTEALLTHWPGVCHATVNDVFLSTFALAVAGWRGRGEDADAPV	3615
QY	82	-----	81
Db	3616	VLDLSEHREHAAVPGVELSRGTGFTSMYPVRLAPPAGAGSGSALRALKAVKEQLTV	3675
QY	82	-----	81

Db	3676	PGDGLGYGLLRHLNPRTRAALAAALPLPEFGFNYLGRIGOBGTDEAPWTIEGGDVAGIDGA	3735
QY	82	-----	81
Db	3736	MPLAHPVDVNAVARETADGTRLRARWTYSRTALEPEDTORLADTWFRLLRLVEEAROPG	3795
QY	82	-----	81
Db	3796	AGGLTPSDIAHPALAQDEIEDLEHTVPGLODILPLAPLQSGFLFLNLYDENARDVYVQL	3855
QY	82	-----	81
Db	3856	AFDLEGSDGTRMRAAGALLRRHANLRAGFRQTATGTWVQVPAELEPDWRCDLTDR	3915
QY	82	-----CWSQ- 	85
Db	3916	DEARDAEAGRLAAGDRERRFDLTSPPLMFTAIRLSADRVLVMTNHHILLDQSWPPLL	3975
QY	86	-----	85
Db	3976	WOELTELYVSGGDPVSLPPVPRPYRDHLAWLGARDRDAARDNRSLSGLDEATLLAPDAG	4035
QY	86	-----	85
Db	4036	PAEAAPLGIPGLDRDATAALSAAWARGVTMTNTVQCAWALAAQATGRDDVVFATVS	4095
QY	86	-----	85
Db	4096	CRPPELPGVESMIGLFINLTPVRARLDOABPLGDLFRRLONEQARLLDHOHQPGLADIQH	4155
QY	86	-----	85
Db	4156	ACHGELFDTAMVFQNPVSADTTSRQLDGLRVAGYDAVESTDPVNLVAHTRDDALRL	4215
QY	86	-----	85
Db	4216	DYRADACAGDLVRSADRMLRLEALVTDSRPAVHLDTLDPVREVRVLVEMGAPTOLP	4275
QY	86	-----	85
Db	4276	GTPLHELISEOARLTPDAVAVVCDGTSITYAELGGANLARHLLGEGLAGEDFVAIALA	4335
QY	86	-----	85
Db	4336	KSLDAVISMLAVLKTGAAYLPIDPDYPAERITYMLDDARPALTLTEPVVERYTGHSVTA	4395
QY	86	-----	85
Db	4396	VTDEERRSPWSARHAAVMIYTSGSTGRPKGVVIEHHALATYLHRAPTYTAMTGVTVLHS	4455
QY	86	-----	85
Db	4456	PLAFDLTITALWTPLTSGTVHLTSLESDTOPSLIKATPSHLPLLTLPETASPSHTLI	4515
QY	86	-----	85
Db	4516	LGGEALHTDHLVTWTHQHVQIINAYGPTESTVNIIDHVGEDTPDGPVPIGRPFANTQ	4575
QY	86	-----TINPE- 	90
Db	4576	VYVLDALRPVAPGVGTGELYLAGEQLARGLPALTAERTFANPHSSTPGARMYRTGDL	4635
QY	91	-----	90
Db	4636	AHWNHHGLTYDGRADHQIKLRGHRIEPEIEATLTAQTGITOATVOLBEDQDORLVA	4695
QY	91	-----	90
Db	4696	YLVVNDSTEYDEKTVRDALTSALPDYMVPSALVTLDALPLTPNGKLDRTALPAPAYSAST	4755
QY	91	-----	90

Db 4756 AGRAPRTPREVLCTLFAEVLGVDLVTIDDNFFDLGGHSLLATRLVSRTRTALGVLSIR 4815
QY 91 ----- 90
Db 4816 QLFETPTVAGLAELDASGTVRTALTARPRERIPLSYAQORLWFLHQLGSPSATYNTVL 4875
QY 91 ----- 90
Db 4876 TLRGGALDVALRAAISDVVARHESLRTVFTEDERGAYQIVLPVEAASPTFTVDVAEE 4935
QY 91 ----- 90
Db 4936 EIGRLDEAVGHCFDLAQELPARTSLFRVSEREHVHLLLLIHIIASDAWSRAPLAQDLTAA 4995
QY 91 ----- 94
Db 4996 YAAVRSEAPWAPLTVQYADYALWQOEIIGDDTDADSLAGRLAYWKQQLAGLPEQDL 5055
QY 95 ----- 94
Db 5056 PDRPRPAVAGYSGDRVPFTVPTLHTRLTTELARATNTSAFMVIQAAVAVLLTRLGAGED 5115
QY 95 ----- 94
Db 5116 IPICTPVAGRTDAAADDLIGLINTLVLTDTSGDPTFRLLDRVRDTLAAVAHQDLPF 5175
QY 95 ----- 94
Db 5176 ERLVEALNPARTLSHHPLFQVLLTFNNTDHEGALKDISELPLGLTVALREVORTSSKFDLS 5235
QY 95 ----- 94
Db 5236 FGFAESFDSRRPQOIEAALDFSTELLDRRSAQAIADRFLRLVLEAVTTAPDRPIGAVELM 5295
QY 95 ----- 94
Db 5296 DPAERERVLVWNGAPTQLPCTPLHELISEQARLTPOAVAVVCDGTTLTYAELDRRANQL 5355
QY 95 ----- 94
Db 5356 ARHLLGEGLAEDPVAIALAKSLDAVISMLAVLTKGAAYLPIDPDYPAERITYMLDDAQP 5415
QY 95 ----- 94
Db 5416 ALTLTAPIPPASYSRPTSEITDVERSSPWSARHAAYIYTSSTGRPKGVVIEHHALAT 5475
QY 95 ----- 94
Db 5476 YLHRARNTYTAMTGVTLHSPAFDLTITALTWPLTAGTGHVLTSLBEEAEVQPSLIKATP 5535
QY 95 ----- 94
Db 5536 SHLPLLTLPETASPSHTLILGGEALHTDLATWRTQHPGAQIINAYGPTSTVNITDHH 5595
QY 95 ----- 94
Db 5596 VSEDTPDGPVPIGRPFANTQVVLDSALRPVAPGVTCGELYLAGELARGYLGRPALTAER 5655
QY 95 ----- 94
Db 5656 FTANPHSSTPGARMYRTGDLAHNNHGHLYTDGRADHQIKLRGHRIBPGEIETTLTAQTG 5715
QY 95 ----- 94
Db 5716 ITQTTVOLREDTPGQORLVAVLVNDSTEXDEPTLRDALASALPDYMRPSAYVTLDALPL 5775
QY 95 ----- 94
Db 5776 TPNGKLDRTALPAPAYSASTTGRTPRPREIILCTLFAEVLGVDLVTIDDNFFDLGGHSL 5835
QY 95 ----- 94
Db 5836 LATRLVSRARTALGVLSVRQFFETPTIAGLSGAFDRAGRARAALTARPRERIPLSYAQ 5895

QY 95 ----- 94
Db 5896 QRLWFLHQLGSPSATYNIPTTLRLTGTDLTDALQSALNDLLARHESLRTTYTDEGEGPRO 5955
QY 95 ----- 94
Db 5956 VIHAMEPGMLPLGVVDTEGEGELDAMLSAGVHHAFDLTAGIPVRATLFRISEQEHVLLLLII 6015
QY 95 ----- 94
Db 6016 HHIATDAWSRTPLCHDLAAAYSARCAGDYPAMEPLPVQYADYALWQREVLDGEGDADAPA 6075
QY 95 ----- 94
Db 6076 GRQLAYWTRQLADLPEQLDLPDRPRPAVASQDGRVAFSLDADLYVRLTELARATHSS 6135
QY 95 ----- 94
Db 6136 FMVQAAALAVLLTRLGAGEDIPIGTVPAGRTDDATENLVGFFVNTLVLRNDTSGNPTFRE 6195
QY 95 ----- 94
Db 6196 LLEETRTDLAAVAHQDLPFERLVEALNPARTLAHHPLFQVMLILSTAETDPDASLALPG 6255
QY 95 ----- 94
Db 6256 LRVGAERSRLGAAKVDLAFALAEVRDGEGRSTGLTGALDRTDLFORSTARSILVERFVRT 6315
QY 95 ----- 94
Db 6316 LEAVVADPGVRLSRVPLTGSERRSLDRGTGPLEGLDATLPELFAEQALRTPGAPALV 6375
QY 95 ----- 94
Db 6376 RGGTTVSYAELDLRTHRLARLLRQOQVRPGTPVMMLMERSPAHVATLAIAKAGAYVPL 6435
QY 95 ----- 94
Db 6436 HDITYPLDRMRHVADTAATLILTDRAEAAAGQLGARVMVWDEFGAAPSGSEADAAPGTG 6495
QY 95 ----- 94
Db 6496 TGTGTSGRSGYDDAPEVGLRPQDLAYVMYTSGSTGVPGVAVTHRGVVDLVRDHCHWRPG 6555
QY 95 ----- 99
Db 6556 VHERVLLHAPHAPDVSCYEMWVPLVSGGTVVVAPPCHLDPAAITDLITAHDLITAG 6615
QY 100 ----- 102
Db 6616 FFRVAAEAEPCFAGVREVLTCGGDVVSPAARVARLAHHPRIVLRLHLYGPTETTLTCLVTOHE 6675
QY 103 ----- 102
Db 6676 VTAPYEARGSLPVGRATGNTRAYVLDLYLPQVPAGVPGELFISGGLARGYLDRPDLTCE 6735
QY 103 ----- 106
Db 6736 RFVADPYGSGGERMYRTGDLVRYNAAAGELEYLARADDQVKIRGFRVELGELIEAVLATRPE 6795
QY 107 ----- 106
Db 6796 LAQAAVVVREDRPGDRLRGVYVVAAGRDEVPDALRAFROALPDYVMVPSAFVVLGTL 6855
QY 107 ----- 109
Db 6856 PLTANGKLDRLKALPAPDYGAASTGRAARTPABELLCTLFAQVGLGSAGVDDGFFDLGGD 6915
QY 110 ----- 109
Db 6916 SILSIQVSRARAAGLALAVRDVFEHQSTARLAAALTDRODAASVPEAEAVPPYGPAPLT 6975

Qy	110	-----ELPDGNQVR	119	Db	206	QERKRSTASPRSSRGSPSPKSMKSRCTPKRTLKPREGSPSKLRSRTSTPVNE	265
Db	6976	PVMARIAELGLGDDFNQSVVSLPPAVDRDLRALQVLDHDLRLVLPDSTEV	7035	Qy	9	-----	8
Qy	120	G-----VTOLGG-----AC-----	128	Db	266	EVSQSESRSSRTDKMEVDQVSGASKRKPDLPPPGDDEKKLRAGSPSTRKSASASP	325
Db	7036	APGSVPADVLVSVTRAPGATGEARDALLVEAACARDRLAPAEGRMLQAVLVDGTD	7095	Qy	9	-----	8
Qy	129	-----SPTWSCSLITE-----	138	Db	326	TPSRKSSAGGAASGTTGASASATSATSGSSASSDASRDKYTRPPIVLEASRSOTGRIGG	385
Db	7096	GTGGTSGADGVLIIVAHHLVDSVTWISWPDLAAYRGEEPAVGTGWSRWQWATSLARLA	7155	Qy	9	-----	8
Qy	139	-----DTGFDL-----	144	Db	386	SVLEVQWQCHSSTIIIEWYRDGTLVRNSSEYSQSPNGSIAKLQVNLKTEKSGLYKCHAK	445
Db	7156	TDPRVEATAHWEHTLTGAGTLRLDRGRDLOGDAGRISLDLAPHTTEALLTRLPGGVNAS	7215	Qy	9	-----	8
Qy	145	-----	144	Db	446	CDYGEQSSAMVKIEOSDVEEELMKHKOAEDEYKQEEQKSOQLQAEKTKRVARRSKSKS	505
Db	7216	VHDVLLTAFANAVAGWRGCGEDDPAPVVLDLSEHGRHEEAVPGAELSRTAGWFTALHPV	7275	Qy	9	-----ELL	11
Qy	145	-----	144	Db	506	KSPAPQAKKSTTSSEGRQEAASEVHKRSSVRPDPDEESQDDEIPSSGLTIPEERRRELL	565
Db	7276	RLAPDVTWARLHQDQDALRDGLKQVKEQLRSVPGDGLGHLRLHNPNTAGPRLARLPEP	7335	Qy	12	VAVG-----PDVF-----	19
Qy	145	-----	144	Db	566	GQGESDDEVSEISLPSFAGCKPRRKTDFSCVLLYLQSPKQDDMFSRDTLLRKT	625
Db	7336	DGFNYLGRVTPATGTPEPMTVTGGLAASRPTAPMAHVELSAVVHEGADGPRRLRAEW	7395	Qy	20	-----	19
Qy	145	-----GVT	147	Db	626	STKNESSTVEEKTLRKTVKVDGELDFKAMVKLKVKEEGGTTEKSGPLOHADSTSS	685
Db	7396	TYARLVPDHDARRLAEQWFRALAEALVEQADRGGLT	7434	Qy	20	-----	19
RESULT 9							
h27935							
Hypochemical protein ZK617.1b - Caenorhabditis elegans							
C:Species: Caenorhabditis elegans							
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000							
C:Accession: T27935; T28031							
R:White, S.							
submitted to the EMBL Data Library, May 1996							
A:Reference number: 220442							
A:Accession: T27935							
A:Status: preliminary; translated from GB/EMBL/DBJ							
A:Molecule type: DNA							
A:Residues: 1-7160 <MIL>							
A:Cross-references: EMBL:Z73897; PIDN:CAA98065.1; GSPDB:GNO0022; CESP:ZK617.1b							
A:Experimental source: clone ZK617							
R:Haris, B.							
submitted to the EMBL Data Library, May 1996							
A:Reference number: 220458							
A:Accession: T28031							
A:Status: preliminary; translated from GB/EMBL/DBJ							
A:Molecule type: DNA							
A:Residues: 1-7160 <W12>							
A:Cross-references: EMBL:Z73899; PIDN:CAA98082.1; GSPDB:GNO0022; CESP:ZK617.1b							
A:Experimental source: clone ZK629							
C:Genetics:							
A:Gene: CESP:ZK617.1b							
A:Map position: 4							
A:Introns: 10/3; 61/3; 135/2; 168/3; 256/2; 379/3; 405/2; 463/1; 508/3; 542/3; 574/3; 59							
3067/1; 3147/3; 3269/1; 6473/3; 7012/3; 7097/1; 7129/3							
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;							
Query Match 29.1%; Score 225; DB 2; Length 7160;							
Best Local Similarity 1.2%; Pred. No. 1.2e+02;							
Matches 83; Conservative 29; Mismatches 34; Indels 6500; Gaps 26;							
Qy	3	GGTLHL-----	8	Qy	24	-----EDTERY-----	29
Db	146	GGLYHAIFSDLGDTYLCOLEIRGSSDAGQVRCNIRNDOGETNANLALNFEPPSPSR	205	Qy	30	-----VLT-----	32
Qy	9	-----	8	Qy	33	-----	32

Db 1286 EVEDDOAEWTARITQDVSFKVQVVEEPRHTFVVPKMSQKWNESDLATLETVDNDKDAE 1345
Qy 33 -----
Db 1346 VVWHGDKRIDIGVKFKVNESSNRKRLLIINGARIEDHGEYKCTTKDDRTMAQLIVDAKN 1405
Qy 33 -----NLNI 36
Db 1406 KFIVALKDETEIEXDDVTLMCQTKDTKTPGIWFRNGKQISSMPGGKFETOSRNGHTLKI 1465
Qy 37 G-----37
Db 1466 GKIEWNEADYVEIDQAGLRGSCNVTVLEAEKRPILNKKPKKIEAKAGEPCVVKVPOIKG 1525
Qy 38 -----AELLR-----42
Db 1526 TRRGDPRAQLKNGKPIDEENRKLVEVIIKDDVAEIVFKNPOLADTGKWALELGNAGTA 1585
Qy 43 -----DPSLGA-----48
Db 1586 LAPPFLVKOKPKPKGPLETKNVTABGLDLVWGTPDPDEGAPVKAVIIEMQEGRSGNWA 1645
Qy 49 -----48
Db 1646 KVGETKGTDFKVKDLKEHGEYKFRVKALNECGLSDPLTGESVLAKNPYGVPGPKPNMDAI 1705
Qy 49 -----48
Db 1706 DVDKDHCTLAWPEEEDGGAPITGYIIERREKSEKDHQVQOTKPDCCELTDKKVVEDKE 1765
Qy 49 -----48
Db 1766 YLYRVKAVNAGPGDPCDHGKPIKWKAKASPEFTGGIGKDLRLKVGETIKYDVPISGEP 1825
Qy 49 -----48
Db 1826 LPECLWVNGKPLKAVGRVWMSERGHMKIENAVRADSGKFTITLKNSSGSCDSTATV 1885
Qy 49 -----48
Db 1886 TVVCRPTPPKGLDIADVCADGATLSWNPDDGDDPLTGYVEAQDMNDKGVIEVGKV 1945
Qy 49 -----48
Db 1946 DPNTTTLKVLNGLRNGNYKFRKAVNNEGESEPLSADQYQIKDPWDEPGKGRPEITDF 2005
Qy 49 -----48
Db 2006 DADRIDIABEPHKKGGAPIEYIYEVDPDTKEWKEKRVDPDTNASISGLKEGEYQFR 2065
Qy 49 -----48
Db 2066 VRVAVKAGPQGPSEPEKQLAKPKFIPAWLKHDLNLSITVKAGATVRWEVKIGGEPIPEV 2125
Qy 49 -----48
Db 2126 KWFKNQOLENGIQLTIDTRKNEHTILCIPSAMRSDVGEYRLTVKNSHGADAEKANITVL 2185
Qy 49 -----48
Db 2186 DRPSKNGPLEVSDVFEDNLNLSWKPPDDGGEPYIYEVVEKLDATATGRWVPCAKVKDKT 2245
Qy 49 -----48
Db 2246 AHIDGLKKGQTYQFRVKAVNKEGASDALSTDKOTKAKNPYDEPGKTGTPTDVVDWADRVS 2305
Qy 49 -----QFRV-----52
Db 2306 LEWEPKSDGAPITQYVIEKKGHGRDQWQCGKVGSDQTNAEILGLKEGEYQFRYKAV 2365
Qy 53 -----52
Db 2366 NKAGGEASDPSRKVAVKPRNLKPWIDREAMKTITIKVGNDEPDPVVRGPEPPPKKEWIF 2425

Qy 53 -----HLVQMVL-----60
Db 2426 NEKPVDDQKIRIESEDYKTRFVLRGATRKHAGLYTLTATNASGDKHSVEIVLGPSSP 2485
Qy 61 -----60
Db 2486 LGPLEVSNVYEDRADLEWKVPEDDGGAPIDHVIEKMDLATGRWVPCGRSETTKTTVPNL 2545
Qy 61 -----60
Db 2546 QPCHEYKFRVAVNKEGESDPLTTNTAILAKNPYEVPGKVDKPELVDMDKHVDLAWNAP 2605
Qy 61 -----60
Db 2606 DGGAPIEAFVIEKDKNGRWEALVVPDQKTATVPNLKEGEYQFRISARNKAGTGDP 2665
Qy 61 -----60
Db 2666 SDPSDRVAVKPRNLAPRIHREDLSDTTVKVGATLKFIVHIDGEPAPDVTWMSFNGKIGES 2725
Qy 61 -----TEPEG-----65
Db 2726 KQIENEPVISRPALPKALRKQSGKYTITATNINGTDSVTINIKVKSCKTPKGPPIEVD 2785
Qy 66 -----65
Db 2786 VPEDRATLDWKPPEDDGGEPBIEFYIEKMTKOGIWWPCGRSGDTHFTVDSLNLKGDHYKF 2845
Qy 66 -----65
Db 2846 RVKAVNSEGSDPLETETDILAKNPPDRPDRPORPEPTDWDSDHVDLKWDPPLSDGAPI 2905
Qy 66 -----65
Db 2906 EYQIEKRTKYGRWEPAITVPGQTTATVPDLTPNEEYEFRRVAVNKGSPSPSDASKAV 2965
Qy 66 -----65
Db 2966 IAKPRNLKPHIDRALKNLTIKAGOSISFDPVPSGEPAPTVTWHWPDNREINRGVRKLD 3025
Qy 66 -----65
Db 3026 NPEYQSKLVVQMERDGSCTFTIKAVNAGEDEATVKINVIDKPTSPNGPLDVSVDHGDH 3085
Qy 66 -----65
Db 3086 VTLNWRAPDDGGIPIENIVIEKYDTASGRWVPAKVAGDKTTAVVDGLIPGHEYKFRVA 3145
Qy 66 -----APNIT-----70
Db 3146 AVNAEGSDPLETFTGTTLAKDPFDKPKGTNAPEITDMDKHVDLEWKPPANDGGAPIEY 3205
Qy 71 -----70
Db 3206 VVEMKDEFSPPFNDVAHVAPAGOTNATVGNLKEGKYEFIRAKNKGAGLGPSPDSASAVAK 3265
Qy 71 -----70
Db 3266 ARNVPPVIDRNSIQEIKVKAGODFSLNIPVSGEPTTITWTEGTPVESDDRMKLNEDG 3325
Qy 71 -----70
Db 3326 KTKFHVKRALRSDTGYTIIKAENENGTDTAEVKVTYVLDHPSSPRGPLDVNTNIVKDCDLA 3385
Qy 71 -----70
Db 3386 WKPEPDGGAIEISHYVIEKQDAATGRWTACGSKOTNFHVDLDTQGHYKFRVKAVNRHG 3445
Qy 71 -----70
Db 3446 DSDPLEAREAILAKOPDFDRADKPGTPEIVDMDKHADLKWTPPADGGAPIEGYLVEMR 3505


```

Qy 71 ----- 70
Db 3506 PSGDWPAVTVGAGELTATVDGLKPGQTYQFRVKALNKAGESTPSDPRTWVAKPRHLAP 3565
Qy 71 ----- 70
Db 3566 KINRDMFVAQRVKAGQTLNFDVNVVEGEPAPKIEWFLNGSLSSGGNTHIDNNTDNNTKLT 3625
Qy 71 ----- 70
Db 3626 TKSTARADSGKYIVATNESGKDEHEVDNLDIFCAPEGPLRHKDIITKESVVLKWDDEPL 3685
Qy 71 -----ANLT 74
Db 3686 DDGSPITNYVVEKOEDGGRWVPCGETSDTSLKYNKLSEGHEYKFRVKAVNRQGTSAFLT 3745
Qy 75 S----- 75
Db 3746 SDHAIVAKNPPDEPDAPTVDVNDKDHVDLEWKPPANDGGCAPIDAVIVEKKDKFGDWV 3805
Qy 76 ----- 75
Db 3806 ECARVDGTTKATADNLTPGETYQFRVKAVNKGPKGSDPTGNVYVAKPRRMAPKLNLAG 3865
Qy 76 ----- 75
Db 3866 LLDLRKAGTPIKLDIAFEGEPAPVAKWKANDATIDTGARADVNTPTSSAIHIFSARVG 3925
Qy 76 -----SLLSVCQ----- 82
Db 3926 DTGVYKIIIVENEHKGDTAOCNVTLDVPGTPEGPLKIDEIHKEGCTLNWKPPDPTDNGGTDV 3985
Qy 83 ----- 82
Db 3986 LHYIVEKMDTSRGTWGCVGTGCTAKVNKLVPKGEYAFRVKAVNLOGESKPLEAEEPII 4045
Qy 83 -----WSQTIN----- 88
Db 4046 AKNQFDVDPVDPKPEVTDWKDRIDIKMNPNTANNGGAPVTGTVIVEKKEGSAIWEAGKT 4105
Qy 89 -----PEDDTPD----- 95
Db 4106 PGTTSADNLKPGVEYEFVRIAVNAAGPSDPTDPTDQITKARYLKPILTASRKIKKA 4165
Qy 96 ----- 95
Db 4166 GFTHNLEVDFIGAPOPTATWVGDSGAALAPELLVDKASTTSIFPPSAKRADSGNYKLG 4225
Qy 96 ----- 95
Db 4226 VKNELGEDEAIFEVIVODRPSAPEGPLEVSDVTKDSVNLWKPDKDDGGAETSNVYVEKR 4285
Qy 96 ----- 95
Db 4286 DTKNTWVPVSAFVTGTSITVPLKTEGHEYEFVRIVMAENTFGRSDSLNTPDEPVLAKDPFGT 4345
Qy 96 ----- 95
Db 4346 PGKPGRPEIVTDNDHIDIKWDPDRNGGSPVDHYDIERKDAKTGRWIKVNTSPVOGTAF 4405
Qy 96 ----- 95
Db 4406 SDRVQKHTYEVYRVVAVNKGQPSDSSAAATAPKMHEAPKFDLDDCKEFRVYKAGEP 4465
Qy 96 ----- 95
Db 4466 LVITIFTASQPDI SWTEGKPLAGVETDTSQTKLVIPSTRSDSGPVKIKAVNPYGE 4525
Qy 96 ----- 95
Db 4526 AEANIKITVDKPGAPENITYPAVSRHTCTLNWDAPKDDGGAIEAGYKIEYQEVGSQIWD 4585
Qy 96 ----- 95

```

```

Db 4586 KVPCLISGTAYTVRGLEHQOYFRIRAEANAVGLSDYCOGVPVVIKDPFPPGAPSTPEI 4645
Qy 96 ----- 95
Db 4646 TGYDTNOVSLAMNPPRDDGGSPILGYVVERFEKRGGDWAPVKMPVMVKGTECIVPGLHEN 4705
Qy 96 -----CH----- 97
Db 4706 ETYQFRVRAVNAAGHGSPNSGSEPVTCRPYVEKPGADAPRVGKITKNSAELTNRPLRD 4765
Qy 98 ----- 97
Db 4766 GGAPIDGYIVEKKKLGDNWTRCNDKPVRODTAFEVKNLGEKEEYEFVRIAVNSAGEPS 4825
Qy 98 --ADLVI----- 102
Db 4826 KPSDLVILIESQGRPIPDINNLDITVRAGETIQIRIPYAGGNPKPIIDLFGNSPIPEN 4885
Qy 103 ----- 102
Db 4886 ERTVVDVNPGEIVITTTGSKSDAGPYKISATNKYKDKTCKLNVFVLDAPGKPTGPIRAT 4945
Qy 103 ----- 102
Db 4946 DIQADAMTSLWRPPKNGDAITNYVVEKRTPGGDMWTVGHVPGTTLRVRNLDANTPYEF 5005
Qy 103 ----- 102
Db 5006 RVRAENQYVGCEPLETDDAIVAKNPPDTPGAPGPEAVETSEEAITLOWTRPTSDGGAPI 5065
Qy 103 ----- 102
Db 5066 QGYVIEKREVGSTWTKAAGNILDTKHRVTGLTPKKTTEFRVAAVNAAGOGESVNSVP 5125
Qy 103 ----- 102
Db 5126 ITADNAPTRPKINMGMLTRDILAYAGERAKILVPFAASPAPKVTFSGENKISPTDPRVK 5185
Qy 103 ----- 102
Db 5186 VEYSDFLATLTIEKSELTDGGLYFVELENSOGSDSASIRLKVVDPKASPQHIRVEDIAPD 5245
Qy 103 ----- 102
Db 5246 CCTLYWMPSSDGGSPITNYIVEKLDLRHSDGKWEKVSFVRNLNNTVGGLIKDNRYRFR 5305
Qy 103 -----YITRFDLELP----- 112
Db 5306 VRAETQVGSBPCELADVVAKYQFEVNPNOPEAPTVRDKDSTWAELEWDRDRDGGSKIIG 5365
Qy 113 ----- 112
Db 5366 YQVQYRDTSSGRWINAKMDLSEQCHARTGLURONGEFEFRIIKNNAAGFSKPPSPERCQ 5425
Qy 113 -----DGNROVRG----- 120
Db 5426 LKSRFGPPGPIHVGAKSIGRNHCTITWMAPLEDGGSKIITGVNVEIREYVSTLWTVASDY 5485
Qy 121 ----- 120
Db 5486 NVREPEFTVDKLREBNDYEFRWAINAAGKGIPLSGPIKIOESGSRPOIIVKXPEDTA 5545
Qy 121 ----- 120
Db 5546 QPYNNRAVFTCEAVGRPEPTARWLNRNGRELPESSRYRFEASDGVYKFTIKVWDIDAGEY 5605
Qy 121 ----- 120
Db 5606 TVEVSNPYGSDTATANLVVQAPPVIEKDVNPNTILPSGDLVRLKIYFSGTAPFRHSLVLRN 5665
Qy 121 ----- 120

```

D	b	5666	BEIDMDHPTIRIVBFDDHILITIPALS	VREAGRYEYTVSNDGSEATTGFWLNV	TGLPEAP	5725
Q	y	121	-----	-----	-----	120
D	b	5726	OGPLHISNIGPSTATLSHRPPVTDGGS	KITSYVVEKRDLSKDEWTVTSNVKDM	YIVTG	5795
Q	y	121	-----	-----	-----	126
D	b	5786	LFENHEYFRVSAQNGENGAPLVSEHP	IIARLPDPPTSPLNLEIVQVGDDYV	TLWSQR	5845
Q	y	127	-----	-----	-----	126
D	b	5846	PLSDGGRLRGYIVKEQBEHEHDFRCN	QNPSPNNVYNPLIDGRKYRYRVFA	NDAGL	5905
Q	y	127	-----	-----	-----	126
D	b	5906	SDLAELOTLFOASGSGEPKIVSPLSD	LNBEVRCVTFEISGSPREYRWF	KGCKEL	5965
Q	y	127	-----	-----	-----	126
D	b	5966	VDTSKYTLINKDKQVLIINDLTSDDA	DEYTCRATNSSGSTRANLR	IKTKPRVFIPPK	6025
Q	y	127	-----	-----	-----	126
D	b	6026	YHGGVEAQGETIELKIPKAYPOGEAR	WTKDGEKIENNSKFSITDDKFAT	LRISNASR	6085
Q	y	127	-----	-----	-----	126
D	b	6086	EDYGEYRVVENSVCDSGTVNVTVAD	VPPEPRPPIENILDEAVILSWKPP	ALDGGSLV	6145
Q	y	127	-----	-----	-----	131
D	b	6146	TNYTIEKEAMGGSWSPCAKSRYTYT	TIEGLRAGQYEFRIIAENKHGQSK	CPCEPTAPVL	6205
Q	y	132	-----	-----	-----	131
D	b	6206	IPGDERKRRGYDVDEQGI	VRKGTVSSNYDNYVFDIWKQY	PQVPEIKHDVLDHYD	6265
Q	y	132	-----	-----	-----	131
D	b	6266	HEELGTGAFGVVHRVTERATGNN	FAAKFVMTPHESDKETVRKEI	QTMVLRHPTLVNLHD	6325
Q	y	132	-----	-----	-----	131
D	b	6326	AFEDDNEMVIYEFMSGGELFEKVA	DEHNKMSDEAVEYMRQVCKGLCH	MHENNVVHLDL	6385
Q	y	132	-----	-----	-----	133
D	b	6386	KPENIMFTTKRSNELKJIDFGLTA	HPKQSVKVTGTGTAFAAPEVAE	GPVGYTDMWS	6445
Q	y	134	-----	-----	-----	133
D	b	6446	VGVLSYILLGLSPFGENDDET	LRNVKSCDWNDDSAFSGISE	GKDFIRKLLADPNT	6505
Q	y	134	-----	-----	-----	133
D	b	6506	RMTIHALEHPLTPGNAPGRDSQ	IPSSRYTKIRDSIKTKYDAN	PEPLPLGRISNSSL	6565
Q	y	134	-----	-----	-----	133
D	b	6566	RKHPQEYSIRDAFWRSEAPRFIV	KPYGTGEGQSANFYCRVIASS	PPVVTWHKDDR	6625
Q	y	134	-----	-----	-----	133
D	b	6626	ELKOSVKYMKRYNGNDYGLTIN	RVKDDKG EYTVRAKNSYGTKE	IVFLNTRHSEPLKF	6685
Q	y	134	-----	-----	-----	137
D	b	6686	EPLEPMKKAPSPPRVEEKKERRS	APFTFHLNRLLIQKHOCKLTC	SQGNPNPTIEMWK	6745
Q	y	138	-----	-----	-----	148
D	b	6746	DGHPVDEDRVQVSRSGVCSLEIF	NARVDDAGTYTVTATNDL	GVDDV	6791

RESULT 10

T31677

bacitracin synthetase 1 - Bacillus licheniformis

C:Species: Bacillus licheniformis

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Nov-2000

C:Accession: T31677

R:Konz, D.; Klens, A.; Schorgendorfer, K.; Marahiel, M.A.

Chem. Biol. 4, 927-937, 1997

A:Title: The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: mo

A:Reference number: Z21058; MUID:98089193; PMID:9427658

A:Accession: T31677

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-5255 <KON>

A:Cross-references: EMBL:AF007865; NID:g4464275; PID:g2982194; PIDN:AAC06346.1

C:Genetics:

C:Gene: bacc

C:Superfamily: acetate-CoA ligase homology; acyl carrier protein homology

C:Keywords: carrier protein

F:83-524/Domain: acetate-CoA ligase homology <ACLI>

F:542-610/Domain: acyl carrier protein homology <ACLP>

F:1134-1561/Domain: acetate-CoA ligase homology <ACLP2>

F:1583-1651/Domain: acyl carrier protein homology <ACLP3>

F:2171-2601/Domain: acetate-CoA ligase homology <ACLP3>

F:2619-2687/Domain: acyl carrier protein homology <ACLP3>

F:3195-3642/Domain: acetate-CoA ligase homology <ACLP4>

F:3662-3729/Domain: acyl carrier protein homology <ACLP4>

F:4712-5151/Domain: acetate-CoA ligase homology <ACLP5>

F:5169-5237/Domain: acyl carrier protein homology <ACLP5>

Query Match 29.0%; Score 224; DB 2: Length 5255;

Best Local Similarity 1.6%; Pred. No. 47;

Matches 77; Conservative 34; Mismatches 36; Indels 4568; Gaps 24;

Qy 2 AGG----- 4

Db 110 AGGAYLPIDETPKDRIAPMLSDTKAAVLLTQGAADGIDCEADIVOLDREASDGSFKEP 169

Qy 5 ----- 4

Db 170 LSSVNSGSDTAYIYTGSGTGPVKVITPHYSVIRVVQNTNYDITEDNVILQSNYSFD 229

Qy 5 ----- 4

Db 230 GSVDFIPGALLNGASLVMIEKEALLNINRLGSAINEKSVMFITTFALFNMIADIVHVDCL 289

Qy 5 ----- 8

Db 290 SNLRKILFGGERASIPHRKVLNHNVRGDKLIHVYGTSTVYATYYPINEIDDEAETIPI 349

Qy 9 ----- 8

Db 350 GSPLANTSVLIMDEAGKLPIGVPGLCTAGDGLSKGYLNREBELTAEKFIHPPIGERL 409

Qy 9 ----- 8

Db 410 YKTGDLAKWLPDGNIEFGRIDHQVKIRGPRIELGEIESRLEMHEDINETIVTVREDEES 469

Qy 9 ----- 8

Db 470 RPYCAVITANREISLDELKGLGEKLPEYMI PAYFVKLDKLP LTKNGKVDKRALPEPDR 529

Qy 9 ----- 21

Db 530 TAGAENEYAPRNETEKLAAVHVQDLHVEKAGIHDFHQAQMGHSLHAMELIAKIKEMN 589

Qy 22 ----- 25

Db 590 VEIPLHOLFKLATIKELSAFIEANHOEDKGDTLVTRAADPENIHEIFPLTGIOIAYLVGR 649

Qy 26 ----- 28

Db 650 DETFEIGGVATNLTVFEADVDLNRFLQTLQKLDRHPILRTIVFENGQTQILEATQRT 709
QY 29 -----|:|----- 28
Db 710 IETQDLRGFTBEEINVRILEQREKMTSKIIDPSVMPLFELKTFMLPGEKKYFFLNVDP LI 769
QY 29 -----VLTNLI----- 36
Db 770 CODSSMKRLIREFKQLYENPGLQPSLEYSFRDYVLASINFKQTSRYOKDQOQYWLKDLH 829
QY 37 -----|:|:|----- 36
Db 830 FPSAPELPLKSDPAHVAKPSFKFSTFLDGHWTNELLKKARHHHLTPTSVLCAAYILA 889
QY 37 ----- 36
Db 890 YMSRQNHFAINTVFNRIPFHPDVKNMIGDFTSLMLDIIHABENMSFWFALNVQDTLL 949
QY 37 -----GAELLR----- 42
Db 950 EALEHRHVDGVDVIRNIAKKNKGNKAVMPIVFTSVLSENPDDPSLVDNIIHFFSTR 1009
QY 43 ----- 42
Db 1010 TSQVYIDNOVYEINGLYITWDYVEQIPEHEVIESMFDQYIAVIQKAVSGEDVSTIQMNE 1069
QY 43 ----- 42
Db 1070 KSRQMSAYNDTDQSFDAKPLHELFTGVQVKGHPDRMALKHDEVMVTYQELDEKSNQVAF 1129
QY 43 -----DP----- 44
Db 1130 LICKGEKGYIGVIGKRSGLTIVNLLAVLKTGAAYIPLDPDYPERKAYIOSKSNCKFF 1189
QY 45 ----- 44
Db 1190 ISHDVYDKEHIERFSKAPVDRKVDLDDMAYVIFTSGTGKPGQVQITHVPQRTILDINE 1249
QY 45 -----SLGAQFRV----- 52
Db 1250 KFNVTQDNMIGISSLCFDSLVDYVFGALSSGASLVIIDDQDVFSLKETAERITWN 1309
QY 53 -----HLVKWVL----- 60
Db 1310 SVPAIMGMTADVDPNLAHNLRLILLSGDWIPLQLPATIKKTFKNAEVISLGGATEGI 1369
QY 61 ----- 60
Db 1370 WSIYPIQKVEEDWKSIPYKGPLANOKIYVLNQNKQLCPVGEVELYIGGAGVASGYIHD 1429
QY 61 ----- 60
Db 1430 QEKTEHSFIHQELGYIYKTGDYGLVEDGYVEFLGRKDSQVKIRGYRVMGEIENTLVS 1489
QY 61 -----TEPGAPNITA----- 71
Db 1490 HOBITKASVIDYTSPOGINKLYAFVVAENAISSOLDVKEFLQKTLPDYMIIPAKFVQIEEP 1549
QY 72 ----- 71
Db 1550 LTVNGKVDKRTLHDLAEQHTADEGQGRMLPENETQMLLEIWKDIFGLDSINLDVSY 1609
QY 72 ----- 71
Db 1610 EIGGDSLKAIISIETINKRMNVEPISEIFKNDITIALDHYLKNREDSMEHPQIAREK 1669
QY 72 ----- 71
Db 1670 EYPTSPAQQMYMLSMLENERGAYHIPMALLVEGRINAMOLENALKTLQRIEILRTGF 1729
QY 72 -----NLTS----- 75
|:|

Db 1730 EIQNNELIQIYENVDRFLREYECLDASITDQHALMEITSRYCKESIKPPDLSRPPLMRK 1789
QY 76 ----- 75
Db 1790 LJ KIDDIRHILVINFHIIISDGVSQGILMNEILELYSNVPLPEVNVQYKDYVENHTPNQ 1849
QY 76 ----- 75
Db 1850 SAMKKQEAAYWLDVYRDIPSKLDFFPYDYKRHHIDTPEGSSVFLMERELSDHIRKLAKHN 1909
QY 76 ----- 75
Db 1910 GTTLYTVMLSAYYVLLNKYTNQTDIVVGTAAGSLHPLDQDVGVFVNTIALRNEVTSY 1969
QY 76 ----- 75
Db 1970 SFKEFLOOTKERTTAAFDNSEYFPDDLIRKLVGVRNSRNPPLFDTMFVLEDAARMFTKQG 2029
QY 76 ----- 75
Db 2030 DVKLSPIIFELDNAKPDMIENVLDFEQKIVLNIEYSTSLFKDETIOKIAEDYPRILEVS 2089
QY 76 ----- 75
Db 2090 ENLDVALHQIDMISROEKRTLLESFNHTKAYPKGKAIHQLFEEQAKRI PDHTAVVPEDQ 2149
QY 76 ----- 75
Db 2150 KLYRQLNEKANQAVARLLREKGVKPDTLVGIMMERSSDMIAAILGVLKAGAYLPIDPEY 2209
QY 76 ----- 75
Db 2210 PPERMYMAFDSEVKVLIISDVPLAEELTAESIELIHMDDERAGDRSDIDNVNQSGLA 2269
QY 76 -----SLLSVCCWSQ----- 85
Db 2270 YVIYTSSTGKPGKMWIEHQSILNLCMSHOSCFEVGQNDNSSIYASISFDAFVWELFPYI 2329
QY 86 ----- 85
Db 2330 TAGATVHVLNOETRLDVEKLNRYPHDHHITISFLPTPVCEQFTALDNHSLRLLTGGDKL 2389
QY 86 ----- 85
Db 2390 NVFKEKSVQIVNNYGPTENTVWATSPIDKSHQNIPIGKPIDNVKVYILNKDQLCLPLGA 2449
QY 86 ----- 85
Db 2450 SGELCIAGEGLARGVNRPELTREKFIGNPFVPGERMRYTGD LAKMLPDGNIQFLGRVDQ 2509
QY 86 ----- 85
Db 2510 QVKIRGYRIEPEIENLLKYKEIEEA AVIAREDCDHPYLCAYTVTVKVEPEKIRAPL 2569
QY 86 -----TIN-----PED----- 91
Db 2570 KKSLPDYMIPOYFVQLDGLPLTVNGKVDKKS LVPERSVTMDRRYEAPRDQMEKLVSIW 2629
QY 92 ----- 91
Db 2630 EEALGINKIGINSHFFEAGGHS LKAAALVSTIHKELNVKLPLRQIFETPTIKGLRDISVR 2689
QY 92 ----- 91
Db 2690 RRKCFYIDRKTEEPYRSLSSAQKRLYILSOTGSHVAYNMPPAMTLEGDFDIRRENTLK 2749
QY 92 ----- 91
Db 2750 NMVKHESFRTSFVMIDGEVMQOIEKEIDFQVAYSDICKESAEEKIKSIFRPFHLEKAPL 2809
QY 92 ----- 91
Db 2810 LRAEVVKNLREHLLMFDMHHIISDGVSTDI FIOELGALYBECKSLKPFHIQYKDYAEWEN 2869

```
Qy 92 ----- 91
Db 2870 SHARSELKROEYMLTKYKGBIPVLDPIDHKRLTKSSEGDVTAAIESETFRKLQHM 2929
Qy 92 ----- 91
Db 2930 AKENGVTMYMLLLAGYTALLSKYTQEDIIIVGTPAAGRHHEDIQHLIGMFVNTLAIRNHP 2989
Qy 92 ----- 91
Db 2990 EGKKTFRDYLOEVKENTLQAYENQDYPPEELVEKVINIKRDMARNPLFDTMLVYHNTDKVP 3049
Qy 92 ----- 91
Db 3050 FEAGLSRLVEIKRGISKFDITWTASEAADGLRLEVEYSTTLFNKERMERLSEHLISLL 3109
Qy 92 ----- 91
Db 3110 EQAADHPDIAINQIDVLTGGERHRVLDNFRTDGVFCKEMTIPELFEKQAEKTPDHPAVA 3169
Qy 92 ----- 91
Db 3170 FGDETISYRELNERANSIAFTLRQKGVCPDVIAGILTERTSIEMIVGIMGILKAGAYLPI 3229
Qy 92 ----- 96
Db 3230 DPAYQERISYIVKSDSVLCAAGDVPGEAYTGDIIIRIDQTGONDHVENLKHDIKPOH 3289
Qy 97 ----- 96
Db 3290 LAYVIYTSGSTGPKPGVMIHHSSVNNLVHGLNERIYOHDLAHLNVALVAPYIFDASVKQI 3349
Qy 97 ----- 110
Db 3350 FAALLFGHTLCIVPRETAMDAMSLIEYYSKNNINVSMDTPAHLNMLAYVDKTELEFDVKE 3409
Qy 111 ----- 110
Db 3410 LIVGGDALTPDIVGGLFHKFPNLSCNITNVYPTCCVDAASHQIESGKVPQTPSIPIGR 3469
Qy 111 ----- 110
Db 3470 PLLNTSIYIVDKELRPLVPVGIAGELCTIAGEGVARGVYNNRPELTAEPVDFHPFGKMYK 3529
Qy 111 ----- 114
Db 3530 TGDLLMWLPDQIEFLGRADHQVKIRGYRIELGEVEQQLLTHEKIKEAAVIAKQDQNGS 3589
Qy 115 ----- 114
Db 3590 YLCAYIASDKELPAADVROFLEREMPDYMIPIYFVKLDRLPRTPSGKVDRSALPEADGNV 3649
Qy 115 ----- 114
Db 3650 NVMEGTGYPDRNEIERKLVQVREILGABDIGISHHFFAAGGDSIKALQIVSRLAKMNL 3709
Qy 115 ----- 124
Db 3710 KLEMKALFANPKIKOLSRFITEETRRHKGKVPVGETELLPIQKRYFANNKEELDHFNOS 3769
Qy 125 ----- 124
Db 3770 FMLFRKGDYDENIVRTAFNKKILEQHDALRMIEBKGDIIQVNRGYRENFLDLDVYDVVRG 3829
Qy 125 ----- 124
Db 3830 FDSQEEKVELATGIOKSSIRKGLVHLGIFRADEGDHLLIAIHLVVDGVSWRILFED 3889
Qy 125 ----- 124
Db 3890 FETLYLOALKEPLDIGYKTDTSYQEFARQLKYYAQSRRLLKEREYQKALEADVFPFAE 3949
```

```
Qy 125 ----- 124
Db 3950 KLERDTEHSATLSIRIGPDVTAKLLRNAPKAYNTEINDILLTALIAAVRDITGENKLKV 4009
Qy 125 ----- 124
Db 4010 MMEGHGRELDIGVDITRTIGWFTTVVPFIDLGESEISQNIKMVKEALRKIPNKGIGY 4069
Qy 125 ----- 124
Db 4070 GVLKYMTEELQKIQTQAPLPSFNFGEMNDNRKVPFSQSPFSGESIGGKIVRHCAIEMN 4129
Qy 125 ----- 131
Db 4130 AISLNGELTIYTTFNQDQYQSTIEQLNOSFKENLEKIVDHCVDKESDWTSPDYGDVSL 4189
Qy 132 ----- 131
Db 4190 GLELELIKDYSAFOIEKIYPLANMOKMLFHNAMDOTSGAYFOOIVIKLGRVHPDIL 4249
Qy 132 ----- 131
Db 4250 EESFHEIVKREILRASFEYEITAEPROIARDRKTPTTSIDLGTENRTRQHRFIETYLK 4309
Qy 132 ----- 133
Db 4310 EDOEKGFDSLSEALMRVCLIKMSDESYRLIWSHHHLLDQWCLGIVLSLFSLYGKIMKG 4369
Qy 134 ----- 134
Db 4370 ESRLKEPKPYGDIYKMLEKQDQBEAVYWKDYLGYESRSELPAPNRGATSEYCGKEK 4429
Qy 135 ----- 134
Db 4430 VISFSKELTTKTRIARKOHVTVINTVLOGIWGMILAKYKNTDEVVFGTVVSOREAPVGOI 4489
Qy 135 ----- 137
Db 4490 EBMVGLFIHTIPTRISFEGARSFKEVLKKTQAESIESNRYSYMNILSEIQVLSEMKRELIT 4549
Qy 138 E-----DTGFDL----- 144
Db 4550 HVMFONVAFDELPFRSOGSETGFELEGVHGKERTNYNPNLTGVLEDBQLKLKLTFNENV 4609
Qy 145 ----- 144
Db 4610 YDNTIETLEKHIIITVAEQVAEDETQTLRDINLVSKEQHRILHTFNDTKGYPKDKPLH 4669
Qy 145 ----- 144
Db 4670 ELFEQAMKTPDHTALVFGAORMTYRELNEKANQATARLLREKIGRGSIAAIIADRSFEM 4729
Qy 145 ----- 144
Db 4730 IIGIILKAGGYLPIDPETPRDRIDYMLKNVSGAALLVTTDSLLKPPDIKTVDLCSDEL 4789
Qy 145 -----GVTI 148
Db 4790 HLLSEENLPRVNRSDTAYIIVYTSGSTGTPKGVVI 4824
```

RESULT 11

T17428
FK506 polyketide synthase - *Streptomyces* sp. (strain MA6548)
C:Species: *Streptomyces* sp.
A:Variety: strain MA6548
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: T17428
R:Motamedi, H.; Shafiee, A.
Eur. J. Biochem. 256, 528-534, 1998
A:Title: The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant rapamycin
A:Reference number: Z18779; MUID:98451508; PMID:9780228
A:Accession: T17428

A>Status: preliminary; translated from GB/EMBL/DBSJ
A.Molecule type: DNA
A.Residues: 1-7576 <MOT>
A.Cross-references: EMBL:AF082100; NID:g3798623; PID:g3798624; PIDN:AAC68815.1
A.Experimental source: strain MA6548
C.Genetics:
A:Gene: fkbB
C.Function:
A.Description: involved in synthesis of the backbone of the immunosuppressant FK506 poly
C.Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase
C.Keywords: carrier protein
F:54-500/Domain: acetate-CoA ligase homology <ACL>
F:1095-1166/Domain: acyl carrier protein homology <ACPI>
F:1204-1599/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F:1682-1953/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F:2680-2751/Domain: acyl carrier protein homology <ACP2>
F:2804-3198/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
F:3295-3569/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F:4320-4391/Domain: acyl carrier protein homology <ACP3>
F:4435-4810/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
F:5903-5974/Domain: acyl carrier protein homology <ACP4>
F:6018-6412/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
F:6513-6785/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
F:7450-7521/Domain: acyl carrier protein homology <ACP5>

Query Match 29.0%; Score 224; DB 2; Length 7576;
Best Local Similarity 1.3%; Pred.No.1.7e+02;
Matches 95; Conservative 17; Mismatches 34; Indels 7287; Gaps 28;

Qy 1 AAGGI----- 5
Db 51 AASGVGRGRVRLLLGNRTTETIEVLAARAAAVAVPLNPDAGDAELAHFLTDSGATVLV 110
Qy 6 LHELLVAVGPD----- 17
Db 111 TDLLHDQVRRVGTATVVLAGRREPGCVAYEDLAGEPPRPRDDGLDEPAMMLYTSG 170
Qy 18 -VFOAHO-----EDTE----- 27
Db 171 TTGRPKGVVSAORSLGWSATYCDVPSWELTDEDELLNPAFLSHLGHLLCLLAVLTVGAS 230
Qy 28 -----RYVL----- 31
Db 231 ARILGFGVARDVLDALAEHPCTVLVGVPTMYRYLLGAVSGEPOTRALRVALVAGSTSPAS 290
Qy 32 ----- 31
Db 291 LTAGFEAAFGVPLDITYGTETGTSGLTANTLTDTTRVPGSCGLPVPGLSLRFVDPVTGADV 350
Qy 32 ----- 31
Db 351 RRGDEGLWASGPSMLGLGHAQTEATAQVLADGHWYRTGDLARQAETHVITGRVKELII 410
Qy 32 ----- 31
Db 411 RGENIHPREIESVAQVPGVRDAAAAGRPVLPVLYVYVPEGVVPADAILAECCR 470
Qy 32 ----- 31
Db 471 QLAYFKVPDEILHVTVPRTASGVKVRGELAGLPAVLVGTSGGEAALCELWVERREL PDT 530
Qy 32 -----TNLNIGA 38
Db 531 VSQVQAVVTRAVGIGPAEVPDADQALRWDEALVRQAAEPGSGFVLVDTDITDITDANNGD 590
Qy 39 E----- 39
Db 591 EADELVAAAAALGEPQVALRSGVAVPRLVRADTTPLPASQWALRPPASGTLRLDAVAPS 650
Qy 40 ----- 39
Db 651 DAPPRPLAAGEVRVDVRAAGLNFRDLVIALGTYPGDGANGGEAAGIIVTEVPGVDDLVPG 710

Qy 40 ----- 39
Db 711 AHVFLVQDAFRRSVADRRRLVARIPRGWSPVAASVPIVFATAWYGLVDAGVVRPGQKV 770
Qy 40 ----- 39
Db 771 LIHAATGGVGSAAROIARHMGADVYATSPAKOHLLLADGDFADHVANSRTTAPADAFPP 830
Qy 40 ----- 39
Db 831 VDVVLSLTGFLDASVELLAPGGRFTIEMGKTDIRHDVQVQPFLLSDVDPARLEILLELL 890
Qy 40 ----- 39
Db 891 ELFRDELRLPVRPWRDIRRARDAFGMMSOARNTGKTVLTVPRPLGPDAPVLVTGEGADA 950
Qy 40 ----- 44
Db 951 VAGLLRDETGRTHVSSSSAAEAEEGEGVNSTLAHLVLLGDPAGADAYARSASEGLP 1010
Qy 45 -----SLGAO----- 49
Db 1011 IIVVHGLPGIEQAALTGELIEAAIEGGGSYVTVRVGSAGARAEAMAGATPILLSGLAGPT 1070
Qy 50 ----- 49
Db 1071 GRAEPGAAGPEWMSRLAAADADRENVLDDLVRDSVATVGLPGAERCAPDRTFRENGLDS 1130
Qy 50 -----FRVHLVQWVI----- 59
Db 1131 LTTVEFANTVAARTGLRVPASTAFDHTPRSFAAHLGASVAAAPAAKALGSGEPVAIVGM 1190
Qy 60 -----LTEPEG----- 65
Db 1191 ACLPQGVTSPEGLWRLVASGTDAITFVDRGWDVDELVDPPDAVGRSTTRHGGFLTE 1250
Qy 66 -----APN----- 68
Db 1251 ATGFDAAFPGISNEALAMDPOORLVLETSWEAFHAGIVPDTLRESDTGVFMGAPHOGY 1310
Qy 69 -----ITANLTS----- 75
Db 1311 GACRDLGLGVTTATQTSVLSGRLSYFYVLOGPAAVTVDTACSSSLVALHQAALRSGECS 1370
Qy 76 ----- 75
Db 1371 LALAGGVTVNATPGSPVFPFSRQRLSPDGRCKAFADAADGTGFAEGVGLVWVERQSDAER 1430
Qy 76 ----- 75
Db 1431 NGHTVLAVVRGSANVODGASNGLSAPNGVAQORVIRQALANAGLSGADVDAVEAHGTGTV 1490
Qy 76 ----- 75
Db 1491 LGDPIEAQVLAASYGOEREVPLLLGSKVSNVGHGTAAAGVAGVIKVMAMRHGLLPKTLH 1550
Qy 76 ----- 75
Db 1551 VDEPSSHVDWSAGAVEVTVETRPWPESGGARRAGVSSFGVSGTNAHVILEGVAESSVRDG 1610
Qy 76 ----- 75
Db 1611 GSSGLVPLPVSAITESSLALQVERLGEYVRGGADAAAVADGLVGRPAVDFRRAVLLGESA 1670
Qy 76 -----SLLSVCG 82
Db 1671 VAGVAVEGARTVFPFGQSGQWVGMGRELWGASEVFAARMRECAAVLEPHTGWDLLDVLG 1730
Qy 83 ----- 82
Db 1731 EAVVDRVEVLOPASHAVAVSLAALWQAHGVVDPDAVVGHSQGEIAAACVAGALSLEDAAR 1790
Qy 83 ----- 82

Db 1791 VVALRSQAIAARLAGRGMAASIAVPASAVETVEGVIAARNGPESTTVACDPAVERVLA 1850
Qy 83 ----- 82
Db 1851 RYAEGRVRRRIAIDYASHTPHVEAIEAQLADALEGITSPTSPWKMSTVDSGMVTEPG 1910
Qy 83 ----- 82
Db 1911 DAYWYRNLROPVAMDTAVSELDSLFIECSAHPVLLPALDOERTVASLRTDDGMDRFLA 1970
Qy 83 -----WSQ----- 85
Db 1971 ALAAMTQGAADVDMWTILIEPAPHRVLDLPTYFDHKRYMLQAPAPVTGAGIGHFFLSSAAV 2030
Qy 86 -----TI----- 87
Db 2031 LPSGDGVLTLGKLSLATHPLWADHTVOEVVLLPGTAFLELVVRAGDEVCGCDIDELVIET 2090
Qy 88 ----- 87
Db 2091 PHSLPASGAVDLTVTVDPDDTGHRPVSVHARPAGTDWTWTRHATGTGLRTEDTAPDTSNF 2150
Qy 88 ----- 87
Db 2151 POWPPTGAQPASLDEFYERLAGAGYTYGPAFOGLRAAMRAGDTVYAEVALDGAADVDRFG 2210
Qy 88 ----- 87
Db 2211 VHPALLDAAHAGRLDAGDELELPSWTCVRLHATGAAAVRVAHSRGPAGVAVQADLDG 2270
Qy 88 -----NPEDD 92
Db 2271 RPVSVDAVLVRPSVAPSGPNLLGLEWLPVAEAHYDGADELPEGYTLITATHPDDDD 2330
Qy 93 ----- 92
Db 2331 PTNPHTPTRHTQTRVRLTALOHLLTNNHTLIVHTTDPGAAVTGLTRTAQNEHPR 2390
Qy 93 ----- 92
Db 2391 IHLIETHHPPLPLTLTLQHLRLTNNTLHTPLHTPITTHNTTTTTPNTPLPN 2450
Qy 93 -----TDP----- 95
Db 2451 HAILITGSGTLAGILARHLNPLTYLLSRTPPTTPTRTHIPCDLTDPTQITQALHIP 2510
Qy 96 ----- 95
Db 2511 QPLTGIFHTAATHDDATLTLNLTPOHLTTLOPKADAAWHLHHTQNOPLTHFVLYSSAAA 2570
Qy 96 ----- 95
Db 2571 TLGSPGOANYAAANAFDALATHRHTQGPATTIANGMHTTTLSLTQDTEROQVRDG 2630
Qy 96 ----- 95
Db 2631 FRPLTEAGNHLFDASLATDAPFVVAIAI PAEPARRVERRTARTGDSGRDLLAVVCVAT 2690
Qy 96 -----GHAD----- 99
Db 2691 AAVLGHADASEIGAATAFKDGLDLSGLRNLRLAEATGVRLSATAVFDHPTPDALAAR 2750
Qy 100 ----- 99
Db 2751 LABELRGDDDSQGAANAAAAAGPTVPTVAAGDEPLAIVAMACRMFGGVDTPEDLWSLVES 2810
Qy 100 ----- 99
Db 2811 GGAITEFTPTDRCWDLAALYDPDDAIGKVSVRHGGFLSGAADFAEFPGISPREALAND 2870
Qy 100 -----LVLYIT----- 105
||| ::

Db 2871 POORLVLEVSWEAFERAGILPASVRGSDAGVFMCAFTQGYGAGVDLGGFGATGTPTSVL 2930
Qy 106 ----- 105
Db 2931 GRLSYFGLGEPSTVTDACSSSLVALHQAAIRSLRSGECSLALVGGVTVMATTTGTFVEFS 2990
Qy 106 ----- 105
Db 2991 RORGLAPDGRAKAPADTADGTSFAEGAGVLIIVERLSDATRLGHPVLAVVRGSAVSDGAS 3050
Qy 106 ----- 105
Db 3051 NGLSAPNGPAQORVIERALDDAGLAPCDIDAVEAHGTGTRGLDPIEAQALEAAYGLDLRH 3110
Qy 106 ----- 105
Db 3111 PLLIGSLKSNFHTQAAAGVAGVIKVLAMRHGVLPRTLHVDEPSRHVDWGDGVDRLLRN 3170
Qy 106 ----- 105
Db 3171 EPMPVTGRVRRAGVSSFGISGANAHVULEAGPPAAPALPAADPVPEDVVMVPSARTPDG 3230
Qy 106 ----- 105
Db 3231 VRDVAGRLTALTAPAAAAIGHSLATRTTAMRRAVVPARDAEAFARGSEVPVGVGTADVT 3290
Qy 106 ----- 105
Db 3291 DTRAVFVFGQSQMDGMAELLATEPVFARRLGECAELAPYTGWDLLDVIAIRRRPGAPE 3350
Qy 106 ----- 105
Db 3351 LORVDVQSPASFAMMVAELWRAHGVAAPAAVGHSGEVAACVAGVLTLDAAKVVAL 3410
Qy 106 ----- 105
Db 3411 RSLVATERAGHGMVSPADFDAAWAGRLEVAANVNGPASIVVAGAADAVEELLAATP 3470
Qy 106 ----- 105
Db 3471 HARRIADVASHAHVESIRGALLDALADLTGAPETPFSTVDEAMLDROPADAAYWYDN 3530
Qy 106 -----R 106
Db 3531 VRCVRFCAAAARLAELCHRVFVEASHPVLTALADTLAGHPNTAVTGLRRGGGARR 3590
Qy 107 F----- 107
Db 3591 FTRSLAELWVRGVPVSNWPFGLRGVPLPTYPFRDRYWDAPAGTSGHPLGLSLVERAD 3650
Qy 108 ----- 107
Db 3651 GEGAFATLFSRRQPLADHEVDGRIIVFGSALVELLAEGARLGTPTIAELTIVAPVV 3710
Qy 108 ----- 107
Db 3711 VDEERDEIQLTVATEVSGRRVRLHSRTGAGPWTGATGALSADTRDPEPVEWPADA 3770
Qy 108 ----- 107
Db 3771 PVDLTGPDYDLPISYGPAPFRAMTAMWAGORAYASVRLGEQLTDARYGLHPVLDAVHA 3830
Qy 108 ----- 107
Db 3831 LGTLFADPERRRLAFSWSGVRIHARAATQORVULLERADGDTIRILATDEHGAPVLDVDGL 3890
Qy 108 ----- 107
Db 3891 TVRAAEPGTALFEVAVSWVPSPVGMTHLADVPEGDVGEGDGEVSDPPPPVVVL 3950
Qy 108 ----- 107
Db 3951 AVEPGDPDSSPGVRARELGRDLLATVTRTLAAAPRWAESRIVVTRTGTGPAQALGVLRT 4010

Qy 108 ----- 107
Db 4011 AGTEHPRVGRIBADEITPATVEAGLAGDESHVRVTDGTVRAARLRRVTASTGASPLSG 4070
Qy 108 ----- 107
Db 4071 GTVLVTGGTGLRLLVDHLLTAHEAEVVVASRGRPGCAPADDRVRYVAADVTDRDEL 4130
Qy 108 ----- DJEL--- 111
Db 4131 AALVDGVAERLCVVMHAGIVDDAVVATMRPQOWDAVLRKVADVAQWLHRLTRDLELAFA 4190
Qy 112 ----- 111
Db 4191 VLYSSISATFGGAGQANYATGNAFLDALARHRRDOGLPAVSLAWGLWDEADQMGRLAAT 4250
Qy 112 ----- 111
Db 4251 DLTRIARGGMI PMTAEQGLELFDAAALHSDRAALVPVRDLAAVAASDQVPPILRDLVPNA 4310
Qy 112 ----- 111
Db 4311 RRAAAPTAAARDMLVRTSAALVLGHRDAHAVEPTRAFKEVGFDSLTGVELRNRLADATG 4370
Qy 112 ----- 111
Db 4371 LTLPATLVFDHTAQALAAHLDELAGARTATRRRTTAPARRHDEPLAIVGMACRLPGGVA 4430
Qy 112 ----- 111
Db 4431 SPDDLWRLLESGGDITAFPADRGWDVLDALYDPDEHSGTSTVRHGGFLTGAAGDFDAGFF 4490
Qy 112 ----- 111
Db 4491 GISPREALAMDPOORVLETSWEALEHAGIDPHTLRGTTGTGVFMGGYFYGYSGADRGDF 4550
Qy 112 ----- 111
Db 4551 GATSTQTSVLSRLSYFYCLEGPVAVTDTACSSSLVALHQOQSRLTGECSLALVGGVTV 4610
Qy 112 ----- PDGN--- 115
Db 4611 MASPSGFVDFSOQGLAPDGRCKAFAEADGTAFAGSGVLVVERLSDAERHGRVLAUV 4670
Qy 116 ----- 115
Db 4671 RGSVAVNQDGASGLSAPNGPSOERVIRQALANAGLQPSVDIAIEAHGTGTRLGDPLEATA 4730
Qy 116 ----- 115
Db 4731 LLATYQDORATPLLGLSLKSNIGHTQAAAGVAGIIMVLMRHRGTLPTLHVDTFSSHVD 4790
Qy 116 ----- 115
Db 4791 WSAGTVELLTDARTWPETGRPHRAAVSFGVSGTNAHVIIESHPRPTPAPDTGSGSTHVPV 4850
Qy 116 ----- 115
Db 4851 LLISARTPEALDEHTTRVRAFLDAEGEDEERAVASALLTRTAFTHRAALIGTDLITGTAEP 4910
Qy 116 ----- 115
Db 4911 DRRLVWLFSGQSORPGMGDELAAYDVFPARTRRDVLALQVPAGLDVHDTGYAQPAVFA 4970
Qy 116 ----- 115
Db 4971 LQVALSAQLDANGVRPDLVGHSGIGELAAAYVAGVMSLDDATELVASARLMQALPPGGA 5030
Qy 116 ----- 115
Db 5031 MAASASERDALPLLCEGVEITAANGPASIVLSGDEDAVLDAARLGRFTRLRTRSHAFHS 5090

Qy 116 ----- ROVR----- 119
Db 5091 ARMEPMLDEFRDVAERLTYHEPKLPMAAGADCATPEYVWVRQVRDTRVFAEOVAAAYDGAAL 5150
Qy 120 ----- 119
Db 5151 LEIGPDRNLARLVLDGIPVLHGEDEARSAMTALARLHTGGVAVDVWPEVIGAAPTDLPULPT 5210
Qy 120 ----- 119
Db 5211 YPPERTRYWLSRAAGDAAPAGOLPVAHPVLTAAMVMPGTGDLVLTGCRVEATDPLATRV 5270
Qy 120 ----- 119
Db 5271 HPGAARAVLLDLRSRRATKRAAAPWRRSPMTPRSALPQSGGLALSVTVAAPGEDGRRAVA 5330
Qy 120 ----- 119
Db 5331 VHARPCTGEWTEHATGILARVARVASAVPETPSPWPPATARPFDPVGLADRLARAGHTDG 5390
Qy 120 ----- 119
Db 5391 PALPRPRAAWADDDAVHAEVALADEQHAERYGLHPALLGAALALAGEGADLPSAFDDV 5450
Qy 120 ----- 119
Db 5451 RVHATGATTVRVAVTATGIIHLADETGSVATVGAVRRRPLTVEGAVPGLLRPDLAETABL 5510
Qy 120 ----- 119
Db 5511 PPTATTGGLLDDPVVPDVILPAHGTGGGPGLETROLDGARVLAALRSFLTDDRIVADAVLA 5570
Qy 120 ----- 119
Db 5571 VHTGPGLPAAAGLVRTAQAEHPRRIVVDVDAEPDTAAPLLAAAAAGLGEPOVVLRGRAY 5630
Qy 120 ----- 119
Db 5631 ARRLTPAVPSGDAPELDPDGTVLITGSGTLAGIVARHLVGHYGVRRLLMLSRGGTASDV 5690
Qy 120 ----- 119
Db 5691 PGAEVTAACDIADROELASALAGIDPAHPLTAVVHTAAVDDGVLTALTPERLDTVLRP 5750
Qy 120 ----- 119
Db 5751 KADGAWHLHELTKDQDLAALCSTPRPPECSATRDRTTPRRTLTLLDELAEOORRAGLPALS 5810
Qy 120 ----- GVTOL----- 124
Db 5811 VAWGLWEPESGLTAGAGDRMRDGTALTSTERGLALLDAALRSPEAALVAADPGGLGSSA 5870
Qy 125 ----- 124
Db 5871 LLRAPHGRPRRAAGASALADRIAGLSAADAEBKAALAVVRECAAAVLGHEGPARIATAT 5930
Qy 125 ----- 124
Db 5931 FKELGVDSLTAVRLRNAFTVRLPATAVDFPTPQAAKLTAAALTGCRVPARARATK 5990
Qy 125 ----- GGACSP----- 130
Db 5991 AVSGHDEPLAIVGMACRLPGGVASPDLLWRLLES GGCGGITAPPADRGWDVLDALYDPDEH 6050
Qy 131 ----- TWSCLE----- 135
Db 6051 SGTSTVRHGGFLTGAAGDFDAGFFGISPREALAMDPOORVVULETSWEALEQAGIVPQTLRG 6110
Qy 136 ----- ITEDT----- 140
Db 6111 SDTGVMGAFSDGYGLGTDLGGFGATGTQTSVLSGRLSFYGLGEPAVTVDTCSSSLVA 6170
Qy 141 ----- 140

Db 6171 LHQAGQSLRTGCSLALVGGVTVMASPGGFVFSQQRGLAPDGRKAFABEADGTAFAG 6230
QY 141 ----- 140
Db 6231 SGVLVWERLSAERHGRHLAVVRGSAVQDNGSLGAPNGPSQERVIROALANAGLQP 6290
QY 141 ----- 140
Db 6291 SDVDAIEAHGTGRLGDIETATALLATYGOHRTPTLLGLSKNIGHTQAAAGVAGIIM 6350
QY 141 ----- 140
Db 6351 VLAMRHGTLPTLHVDTSPSHVDWTTGDELLTOPRDPWQTNRPHRAGISFGVSGTNAHV 6410
QY 141 ----- 140
Db 6411 IIEAHVAEPPALNPGSPVAPGASLFATERPLPVSARTSEAVDGVQVRLREHLAA 6470
QY 141 ----- 140
Db 6471 HPRDDPRVAAALLATRTDFPHRAVLGDDVVGTALTAPRTVFVFGQSQWLGMGREL 6530
QY 141 ----- 140
Db 6531 MAESPVFARMQCADALAEHTGRDLIAMLDDPAVKSRVDVWHPVCWVMVSLAAVWEEA 6590
QY 141 ----- 140
Db 6591 GVRPDVAVHSGEIAAACVAGAISLEDGARLVALRSALLVRELARGAMGSIATAAAR 6650
QY 141 ----- 140
Db 6651 IDGVWVAGRNGTATTIVSGRPDAVETLIADYEARGVWVTLVVDCTPHTFFVDPVDELO 6710
QY 141 ----- 142
Db 6711 RIVAATTSRAPEIPWFSTADERMIDAPLDDEYFNFRNMRNPVGFAAAVAAAREPGDTVFIE 6770
QY 143 ----- 142
Db 6771 VSAHPVLLPAINGTTVGTLRGGAGRVLDLSAKAHTVGVAVDMVTVVAATGAADDAASV 6830
QY 143 ----- 142
Db 6831 TAHDTGTAHDLPTVAFHHERYWIETPATGTDASGLGLDAVDHPLLAASVALPDSDEVLLTG 6890
QY 143 ----- 142
Db 6891 RLSLATHPMLAGHTVDSGEVLLPGPVFVELAGGAEARCDLLDELVTETPLALPATGAVQ 6950
QY 143 ----- 142
Db 6951 VRVTVAAADGARRAVRIHARAHGRTWTRHASGIAAIATTESATADGPPPPARAERIDI 7010
QY 143 ----- 142
Db 7011 ADFYRLDEARTSSAPSSGMSAAWSHGDTLPAEVALDGTLDARDAARYTLHPALLVTALQA 7070
QY 143 ----- 142
Db 7071 GSLAAGSDAGVRLPFAFTGVRVHSSGATKARVTFAAGDGGARVHIADELGPVAEIGSF 7130
QY 143 ----- 142
Db 7131 VTRPLAEADRGDVRLYRRYRTWTVRAPAAGTTAVRYADLGDGATSDATSPDAASPDVVL 7190
QY 143 ----- 142
Db 7191 MRASDPAEVRAALDDPRTDATLVVSAEAGSVAGAVAAALLDAEPGRLLVLTQTDVTPR 7250
QY 143 ----- 142

Db 7251 RAAALARLDEPHRLUTGCRLEAPRLVPAPTTTTTAAPYGDVVLDDGGSBELARRLGEHGAE 7310
QY 143 ----- 142
Db 7311 VIHHKPGKLPETPVSAVVHAGAADSPWELHRLTRDLPLTAFLVLLVPPGCAAGPWEALAE 7370
QY 143 ----- 142
Db 7371 LRRTEGLPAVAFATATEDROTDLDDVACATGEAAVVATAPRRRGRDPSPLMRPVQRPARRSA 7430
QY 143 ----- 146
Db 7431 SDGSLRLRLPDLSPKQEQVLLDVVRDTAATLLGHADARAVTATTAFAKDLGV 7483
RESULT 12
A45086
HC-toxin synthetase - fungus (Cochliobolus carbonum)
C:Species: Cochliobolus carbonum
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 03-Nov-2000
C:Accession: A45086
R:Scott-Craig, J.S.; Panaccione, D.G.; Pocard, J.A.; Walton, J.D.
J. Biol. Chem. 267, 26044-26049, 1992
A:Title: The cyclic peptide synthetase catalyzing HC-toxin production in the filament
A:Reference number: A45086; MUID:93100128; PMID:1281482
A:Accession: A45086
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-5232 <CO>
A:Experimental source: SB111
A:Note: sequence extracted from NCBI backbone (NCBI:120884)
C:Superfamily: Cochliobolus carbonum HC-toxin synthetase; acetate-CoA ligase homolog
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:277-743/Domain: acetate-CoA ligase homolog <ACLI>
F:784-854/Domain: acyl carrier protein homolog <ACP1>
F:1882-2366/Domain: acetate-CoA ligase homolog <ACLI2>
F:2397-2464/Domain: acyl carrier protein homolog <ACP2>
F:3036-3526/Domain: acetyl carrier protein homolog <ACLI3>
F:3549-3618/Domain: acyl carrier protein homolog <ACP3>
F:4186-4655/Domain: acetate-CoA ligase homolog <ACLI4>
F:4680-4750/Domain: acyl carrier protein homolog <ACP4>
F:818,3583,4715/Binding site: phosphopantetheine (Ser) (covalent) #status predicted
Query Match 28.7%; Score 222; DB 2; Length 5232;
Best Local Similarity 1.8%; Pred. NO. 59;
Matches 87; Conservative 20; Mismatches 36; Indels 4702; Gaps 27;
QY 2 AGCI----- 5
Db 304 AGGVCVPIDPRYPVERIRDIIRTNATIALVGAGKTAALFKSADTAVOTINATKDI PHGL 363
QY 6 ----- 5
Db 364 SOTVOSNTKIDDPAPGLFTSGTGVPKCIIVTHSOICTAVQAKDRFGVTSETRVLQFS 423
QY 6 ----- 5
Db 424 SYTFDISIADTFALFYGCTLCIPSEDRMSNLQDYMSVRPNWAVLTPTVSRFLDPGVV 483
QY 6 ----- 5
Db 484 KOFISTLI FTGEASREADTPWIEAGVNLVNYVGAENTLITTRIRKGSNIYG VNV 543
QY 6 ----- 5
Db 544 TRTWTVDSGACLVPGVSGIGELLIESGHLADKYLNRPRDTEAFLDPLWIPNVEGDSVR 603
QY 6 ----- 5
Db 604 RGRFRYRTGDLVRYCDGSLICVGRSDTQIKLAGORVELGDVEAHLQSDPTTSQAAVVPF 663
QY 6 ----- 5

Db 664 RSGPLEARLIALVTGNKDGTFHNOQSLPKPAFAQCPPDLVKYATSSLOQRLPSPVWPSV 723
Qy 6 ----- 5
Db 724 WLGIDFLPMSVSGKLDRAVLQDQLESLSPSDYAEILCTTGLEVDPGGAASSVASDSLDR 783
Qy 6 ----- 5
Db 784 MNDDSLLLTACSRVNLNLPAGKISYSQSFHAGGDSITAMQVSSWKRFTGKRIGVKOLLV 843
Qy 6 ----- 5
Db 844 SPSISTAASCIKSAQDGRNFVAVRPGORI PVSPQIKLFFQTAEASKSWNHYSFLPRI 903
Qy 6 ----- 5
Db 904 DOPIKQTIEDAISLVMPQHPMLQARFERTBEGDWYQIPIIDVERRASVEVIGSLSTDDR 963
Qy 6 ----- 5
Db 964 EAMLRARQSIDLTEGLRQCFNNNVDEASRLFFVVIHHAUVDLVSWRIIMEEHL 1023
Qy 6 ----- 5
Db 1024 ATDSTPDRGEAYQESVPFLAMCQVQAEAVKOIPVDRVTPLPKIPTADFGYWGKLDENV 1083
Qy 6 ----- 7
Db 1084 YGNTVERKIPLCHSITEDLLYKCHDSLTKTIDVLLAAVLVSFRFSFLDRPVPAVFNESH 1143
Qy 8 ----- 7
Db 1144 GREPGEDAVDLRTVGWFTTISPVVPEVSPGIDLVVVRVKDYRWATPNNGDFYSTK 1203
Qy 8 ----- L
Db 1204 YLTOSGILKFEDHLPAILFNIEGRYQAMESEQTVLKPESWHAGEASKDQDPLRRFLCF 1263
Qy 9 ELLVAVGPD----- 17
Db 1264 EISTAVLPDQGLHLTCSWKNRHOGRIRLWLDTLPPAIGEIVSSLASQLTSDVE 1323
Qy 18 ----- 17
Db 1324 LLRLDYSSLDILKKSILSIPAVQTLDDLEGVYVPGPMQDALFLSQSKSQDGAYEVDFTW 1383
Qy 18 ----- VFOAHOED-- 25
Db 1384 RVATSLONSOPAVDIGCLVEAKQDTVALHAALRTVILESSLPATGILHOVLRSHPDIV 1443
Qy 26 ----- TE----- 27
Db 1444 ILDRVDTAAITLDSYPPPTBEGIALIKRPPHRLICTTIEGVLIKFQVNLHVLFDGMS 1503
Qy 28 ----- RYV----- 30
Db 1504 TDKIIQOLSKAVTCRHSNKLPHSESKLHDGTGNRTPKPLAEFIRYIRDPQRKQDSIN 1563
Qy 31 ----- LTNLNG----- 37
Db 1564 YWKNALRGATTCSPFLPDQITSEKAMPQSWASVPIPLCVDSKELSKTLANLIGITMSTM 1623
Qy 38 ----- 37
Db 1624 FQVWAILVIRIYSONQSVFGYLTSGRDAPVDGIDSAVGNFIAMLVCFDDDDGVHTVA 1683
Qy 38 ----- 37
Db 1684 DMARKIHNASNISHOACSLABIQDALGLSTPLFTFTYLPKRPNTNVKAGEPEHHL 1743
Qy 38 ----- AELLR 42
Db 1744 CFBELSMSDPTFELTLFVEPTQESNEVSAHLDFKLSYISQAYATSIATVAHILSELVH 1803

Qy 43 DP----- 44
Db 1804 DPYRALNTLPIVSEHDTAIIIRSNWDLFPATECIHETSRKVVEHPQREAIACSWGSLT 1863
Qy 45 ----- SLGAQFRVHLVKM----- 57
Db 1864 YAELSDLQSRLSHLVSLGIVGTKIPICFEKSMWTVITILAVVQAGGVFVLEPGHPES 1923
Qy 58 ----- VILTEPE----- 64
Db 1924 RLSGIIKQVQAEALLCSPATSRMGALQNIOTMGTFBKIVLEPEFIRSLPLPKPNHQP 1983
Qy 65 ----- GAP----- 67
Db 1984 MVGLNDDLYVFTSGTGVPGKAVATHQAYATGIYEHAVACGMTSLGAPRSLOFASYSF 2043
Qy 68 ----- NITANL----- 73
Db 2044 DASIGDIFTLAVGGCLCIPREEDRNPAGITTFINRYGVTWAGITPSLALHLDPAVPTL 2103
Qy 74 ----- 73
Db 2104 KALCVAGEPLSMVVTVMKRLNLINMYGPTATVACIANOVCTCTTTTSDICRGYRATT 2163
Qy 74 ----- 73
Db 2164 WVQPDNHNLSLPIGAVGELIIEGSLCRGYLNDPERTAEVFIIRSPSWLHDLRPNSTLYK 2223
Qy 74 ----- 73
Db 2224 TGDLYRSADGKIIFIGRKDTQVMNGORFELGEVEHALQQLDPSDGPPIVOLLKRTQS 2283
Qy 74 ----- 73
Db 2284 GEPDLLIAFLVGRANTGTGNSDEIFIATSTSSLSEFTVIKLODQARAMEVLPLMPVP 2343
Qy 74 ----- TSSLLS----- 79
Db 2344 QAYPIEGGIPLTAAGIKDRMLRKLCEPNRNDLISFTSKALSTSVKDAETDTTVDRL 2403
Qy 80 ----- 79
Db 2404 ARIWEKVLGVKVGRESDFSSGSGNSMAIALRAEAQRSFTLFLVADIPTNPLRLADMAKL 2463
Qy 80 ----- 79
Db 2464 FSHQSVSPSSSTLRTKVPISLQKSSGLQTAAPVNSGSPVRRCKENIIDCPVAFEYE 2523
Qy 80 ----- VCG----- 82
Db 2524 EGSDTOLKEASRICGIGSSRSIEDVFPCTPMQEAVALSLIPGAQASYALHAAPFLRPL 2583
Qy 83 ----- 82
Db 2584 DNRFRSAWESTVKAQPIILRSRIISGNSGSSVVVTSATDSIPQLDVSGLDTFLEQQLQVG 2643
Qy 83 ----- WSO----- 85
Db 2644 PAPGAPLRLAFVYSKADDCDYFVISAHAIYDGMNLNIMSQVLALYTNLNGELPPPGPSF 2703
Qy 86 ----- 85
Db 2704 KHFAHNLNLOSCLDSEDFWRKLLVKPDQESFRPDVPVGHKPKATRCTTTFHFPFSMOSK 2763
Qy 86 ----- TIN----- 88
Db 2764 ICTTANTCINAAWAITLAQYSSNKTVNFVGLWGRDFPMIDIEHMTGTPTIVTVPROVNV 2823
Qy 89 ----- 88
Db 2824 PESSVAEFLDLOKSLAVVLPHQHLHRIQALGPBARQACDFSTLLVNNHSSISWSSEL 2883

Qy 89 ----- 88
Db 2884 EAADIVPPLRSSDLIAYPMVVEVENASSDTLDIRVHSDPDCIEVOLLERLMEQFHNQ 2943
Qy 89 ----- 88
Db 2944 TLCRAASFDPKRIAEMLDDTATTHLRTLFSWNSRVKSDPDVAAIAVHKLEETAQSOPA 3003
Qy 89 ----- 88
Db 3004 ESAIVAHGOLSYMOMDRCADVLARQIRKTNMISAQSPFVCIHLLRSATAVWSMLAVLKA 3063
Qy 89 ----- PED- ----- 91
Db 3064 GGAFMPVDISOPRSRLQNLIEESAKLVLTLPESANALATLSGLTKVIPVLSSELVQOIT 3123
Qy 92 ----- DTP- ----- 95
Db 3124 DNTTKDEYCKSGDTPSSPAYLLYTSGTSGKPKGVVMEHRAWLSGFTCHAEYMGFNSCT 3183
Qy 96 ----- 95
Db 3184 RILOFSLMFDLSILEIAVAVYAGCLFIPSDKERVNNLQDFTINDINTVFLTPSIGKL 3243
Qy 96 ----- 95
Db 3244 LNPXDLNISPAGFIGBPMTRSLIDAWTLPGRRLLVNSYGPTEACVLVTAREISPTAPHDK 3303
Qy 96 ----- GHA- ----- 98
Db 3304 PSSNIGHALGANIIVVEPORTALVPIGAVGELCIEAPSLARCYLANPERTEYSPPSVLD 3363
Qy 99 ----- DLVLYIT- ----- 105
Db 3364 NWQTKGTRVYRTGDLVRYASDGTDLFLGRKGOIKLGRGRIELGEIEHHIRLMSDDPR 3423
Qy 106 ----- 105
Db 3424 FHEASVOLYNPATDPRDATVDVQMRPEPYLAGLLVLDVLSLMSRWSMHVNLNANTSNL 3483
Qy 106 ----- 105
Db 3484 QTLVTELKSLRGVLPHYVWPLHFVAVSRPLPTGSSGKLDAFVRACLRELTAPLDGNFPK 3543
Qy 106 ----- 105
Db 3544 VEQVLTINESVLROWGTVLAMPHSIQRGDDFFSLGSSISAMRLVGLARSSGHKLQHE 3603
Qy 106 ----- REDL- ----- 109
Db 3604 DIFMCPRLADWAGQISFVQEASVSPTTSPTIKFDLLDDCEVDEVIDHILPOLDMNKELIE 3663
Qy 110 ----- 109
Db 3664 DIVPCTPLQESLMAATARHGEAYTMIQSIITVLAQAQLKKAMDVVRFDPEVLRTRIALG 3723
Qy 110 ----- 109
Db 3724 PSQOALQVVVKHELSWESFPFISQKDFYRSLGKGLARLAVITQALDTKQPI SHGT 3783
Qy 110 ----- 109
Db 3784 REARTKNSQDTVMVVVGAHHSIYDAHVLSMTIWRRLYREFIGSQADGILEAETSRSEGVVP 3843
Qy 110 ----- ELPD 113
Db 3844 FKSYYEKLRLCKNDSELLFWKEKLRGVSSQFPFASMPRVLEHQPSATQTLITKVSLEPT 3903
Qy 114 GNRQVRGVT- ----- 122
Db 3904 SSRKKGATVATVAYAAWALTIAHYTADPDVVGATLSGRETMAGSISHPESIAIPTIIT 3963
Qy 123 ----- 122

Db 3964 VPLRIIIDFQTVVSDFLSTLQKDIVRAAYFGQWMLNSIAHIDNDCRACGFKSIIVQV 4023
Qy 123 ----- 122
Db 4024 PDEGENHVGRAANPFPOMLESIGHFPAPLVVEVESESTDVILIRMAYPVLVPEKLAHFI 4083
Qy 123 ----- 122
Db 4084 SDTFTTMSNLNANPKAKVESIPALSEAHLAELDVTCPEWILGKAKDEKIRTESHQCLQ 4143
Qy 123 ----- QLG- ----- 125
Db 4144 DLVCRRAQSPNSQAIDSDWGSISYHELDGLSSILAEHLSQLGVRPEAPVCLLFEKSKWA 4203
Qy 126 ----- 125
Db 4204 VVAMIGIIRKAGCFVPLDPSYPHERLEHIISETGSSVITVSAAYSKLCLSLSVRGIVCDG 4263
Qy 126 ----- 125
Db 4264 SVFSSTKKPLPSTADSPSPFSVVRPNQAAVILFTSGTGKPKGVVMEHHSVCSALIALGKR 4323
Qy 126 ----- GAC- ----- 128
Db 4324 MGLGPOSRLQFNYSYWFVMDLLDI FGTLYVGGCLCIPKEEORMNSLGMVQKFKVNTMLL 4383
Qy 129 ----- 128
Db 4384 STSVRLMOPADTPSLETILCTGEAVLQSDVDRWAPKHLHLAGYGTETCIMSUGELTP 4443
Qy 129 ----- 128
Db 4444 SSPANLIGKPVSCQAVVINPLKETELAPYGATGELYIQGTVARGYLHDDVLTSAKAFIVD 4503
Qy 129 ----- 128
Db 4504 POWLTGYKTENOWSRRAYKTGDLVFWGPOSNLYVYRRKSSQVKIRGORVELAEIEEVI 4563
Qy 129 ----- 128
Db 4564 RQHIPPDTVCVLLSSDDQNTRIILGAVLIGDRALGDPEDLEVIGYMDDLKSHIIPAL 4623
Qy 129 ----- SPT- ----- 131
Db 4624 EASLPHHMIPEAYVPFVQLPTLGLSGKLDKTVRRVAGPLAFSLPOASARHPNQPTVHTQ 4683
Qy 132 ----- 131
Db 4684 KLLRQLWCKILPOLDESAVNKQDNFLGIGGDSIAAIKLVALLRQHGISLAWAEIFTRPTL 4743
Qy 132 ----- 131
Db 4744 EAMSSLIDEHNFVSHAGILSDVTRNTSGVMROTTNLIAGRSHMAVEKRECDNSTLPCT 4803
Qy 132 ----- 131
Db 4804 EYQOMFLAGTEAFTGAHSAQFIERLPEKIDLDRLOAFDHCADWYPNLRTQIHKDADTGR 4863
Qy 132 ----- WSC- ----- 134
Db 4864 LLHDISPIGVKVPWSCHYSDDLNTVLSHDKKPPPLGDLPLHRVTIMRHROPTESMLVWTL 4923
Qy 135 ----- LITEDT- ----- 140
Db 4924 NHAAYDAWSLURMLLEHTEAYANPDYEPSYSLGWTAFVLHTENTKEASRSFWSYLSVDK 4983
Qy 141 ----- 140
Db 4984 PARLMFNYNLVSNPQRDLREARINIPKRVLSOATATVLLAGITLLVARVCDTRDVL 5043
Qy 141 ----- 140

Db 5044 HLTGRTLPLAGIENCPOPTITKVPLRIPLMDQDLVTLEDSVAKKITAEMLRVMPPHEHS 5103
Qy 141 -----GFDL 144
Db 5104 GLSAIREFIPOAEGTTTSSGKFHAGSVLGRPLDLVIHPKGGDL 5148
RESULT 13
A88852
protein unc-22 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: A88852
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: A88852
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-6831 <STO>
A:Cross-references: GB:chr_IV; PIDN:CAA98081.1; PID:g3881830; GSPDB:GN00022
C:Genetics:
A:Gene: unc-22
A:Map position: 4
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;
Query Match 28.6%; Score 221; DB 2; Length 6831;
Best Local Similarity 1.3%; Pred.No.1.7e+02;
Matches 81; Conservative 28; Mismatches 32; Indels 6181; Gaps 25;
Qy 3 GGILHL----- 8
Db 146 GGLYHAIFSDLGDTQYLCLEIRGSSSDAGQYRCNIRNDQGETNANLALNFEEPDPSE 205
Qy 9 ----- 8
Db 206 QBRKRSTASPRSPGSRSPSPKMSKREGTPKRTLKPREGSPSKLRSRTSTPVNE 265
Qy 9 ----- 8
Db 266 EVSQSESRSRSTDKMEVDQVSGAKRKPDLPPGDEKELRAGSPSTRKSPRSKASAP 325
Qy 9 ----- 8
Db 326 TSPRKSSAGGAASGTTGASATSATSATSGSSASSDASDKYTRPPIVLEASRSQTGRIGG 385
Qy 9 ----- 8
Db 386 SVVLEVQWOCHSTIIWYRDGTLVRNSSEYSQSFNGSIKLVNKLTEKSGLYKCHAK 445
Qy 9 ----- 8
Db 446 CDYGEQSSAMVKIEOSDVEEELMKHAKDAEYQKEBQKSQLQAEKTRVARRSKSKS 505
Qy 9 ----- ELL 11
Db 506 KSPAPQAKSTTSSEGRQEAASEVEHKRSSSVRPDPDEESQLDEIPSSGLTIPERRRELL 565
Qy 12 VAVG----- 15
Db 566 GQVGESDDEVSISELPSFAGGPRKTDKPKKVSIAPVSTNKSDDPESTPRRSI 625
Qy 16 ----- 15
Db 626 DMRRESVQIELEKTSTPLVPSGASGAPKIVEPVNTVVENETAILTCYVSGSPAPTR 685
Qy 16 ----- 15
Db 686 WFKSREVISGGRFKHTDCKEHTVALALLKCRSQDSBPYTLTIENVHGTDSADVKLLVT 745

Qy 16 ----- 15
Db 746 SDNGLDFRAMLKHRESQAGFKDGGGGGGGEGGKPMTEAERROSLFPGKKVEKDIP 805
Qy 16 -----PD 17
Db 806 LPEKTVOOQVDKI CEWKCTYSRPNKIRWYKDRKEIFSGGLKYKIVIEKNVCTLIINPE 865
Qy 18 VFOAHQEDTERY-----VLT----- 32
Db 866 V-----DDTKYTCEANGVTHAQLTVLEBPMKYSLNPLPNTQEIYRTKQAVLTCKVNT 920
Qy 33 ----- 32
Db 921 PRAPLVWYRGSKAIOEGDPRFIIIEKDAVGRCTLTIKEVEEDQAEWTARITQDVFSKVQV 980
Qy 33 ----- 32
Db 981 VYEPHRFTFVVPMSKQVNESDLATLETVDNDKDAEYVVMHDKRIDIDGVKFKVSSNR 1040
Qy 33 ----- 32
Db 1041 KRLIINGARIEDHGEYKCTTKDDRTMAQLIVDAKNKFI VALKDTVEIKDDVTLMCOTK 1100
Qy 33 -----NLNIG----- 37
Db 1101 DTKTPGIWFRNGKOISSMPGGKPFETOSRNGTHTLKICKIEMNEADYVEIDQAGLRGSCNV 1160
Qy 38 -----AELLR----- 42
Db 1161 TVLEAEKRPILNWKPKKIEAKAGEPCVVKVFPFOIKGTRRGDPKAOILKNGKPIDEEMRKL 1220
Qy 43 ----- 42
Db 1221 VEVIIKDDVAEIVFNQOLADTGKWALELNSAGTALAPFELFVKDKPKPKPLETKNV 1280
Qy 43 -----DPSLGA----- 48
Db 1281 TAEGLDLVWGTPDPDEGAPVKAYIIEQEGRSNWKVGETKGTDFVKVLDLKEHGEYKFR 1340
Qy 49 ----- 48
Db 1341 VKALNECGLDPLTGESVLAKNPYGVPGKPKNMDAIDVDKDHCTLANEPPEDGGAPITG 1400
Qy 49 ----- 48
Db 1401 YIIRERKSEKDWHQVGTQKPDCCELTDKKVVEDKEYLYRVKAVNKAGPGDPCDHGKPIK 1460
Qy 49 ----- 48
Db 1461 MKAKKASPEFTGGGIKDLRLKVGETIKYDVPISGEPLPECLWVNGKPLKAVGRVKNSE 1520
Qy 49 ----- 48
Db 1521 RGHIMKIENAVRADSGKFTITLKNSGSCDSTATVTVVGRPTPPKGPLDIADVADGAT 1580
Qy 49 ----- 48
Db 1581 LSWNPDDDDGDLTGTYVEAQMDNKGKYEIVGKVDPNPTTLKYNGLRNKNGYKFRVKA 1640
Qy 49 ----- 48
Db 1641 VNNEGEPLSADQYTOIKDPWDEPGKPRPEITFDADRIDIAWEPHKGDPGPIEYI 1700
Qy 49 ----- 48
Db 1701 VEVDPDTKEMKEVKRVPDPTNASISGLKEGKEYQFRVAVNKAGPGQSPSEKQLAKPK 1760
Qy 49 ----- 48
Db 1761 FTPAWLKHDNLKSIITVKACATVRWEVKTGGEPFPEVKWPKGNQOLENGIQLTIDTRKNEH 1820
Qy 49 ----- 48

```
Db 1821 TILCIPAMRSVGEYRLTVKNSHGAEKANTVLDRSPKPNGLVSDVFDNLNSW 1880
QY 49 ----- 48
Db 1881 KPDDGGEPIEYVEEKLDTATGRWVPCAKVKTAKAHIDGLKKGQYQFRVAVNKEGA 1940
QY 49 ----- 48
Db 1941 SDALSTDKTKAKNPYDEPGKTGTPVDWDADRVSLEWEPKSDGGAPITQYVIEKKGK 2000
QY 49 -----OFRV----- 52
Db 2001 HGRDWQCGKVGSDQTNAEILGLKEGEYQFRVAVNKAAGPEASDPSRKVAVKPNLKP 2060
QY 53 ----- 52
Db 2061 WIDREAMKTTIKVGNDFEVPVRGEPKKEWIFNEKPVDDOKIRIESDYKTRFVL 2120
QY 53 -----HLVKMVL----- 60
Db 2121 GATRKHAGLYLTATNAGSDKHSVEVILGKPSPLGLEVSNVYEDRADLEWKVPEDD 2180
QY 61 ----- 60
Db 2181 GGAPIDHVEIEKMDLATGRWVPCGRSETTKTTPVNLQPGHEYKFRVAVNKEGESDPLT 2240
QY 61 ----- 60
Db 2241 NTAILAKNPYEPGVKDKPELVMDKOHVLDLWNPADGGAPIEAFVIEKDKNGRWEA 2300
QY 61 ----- 60
Db 2301 LVVPGDOKTATVNLKEGEYQFRISARNKAGTGDPDPSDRVAVKPNLAPRIHREDLS 2360
QY 61 ----- 60
Db 2361 DTTVKVATLKFIVHIDGEPADVTWFSNGKIGESKAQIENEPYISRFALPKALRKQSG 2420
QY 61 -----TPEG----- 65
Db 2421 KYTITATNINGTDSVTINIKVSKXPTKPGPIEVDVFEDRATLDMKPPEDDGGEPYEF 2480
QY 66 ----- 65
Db 2481 EIEKXNTKDGIIWPCGRSGDTHFTVDSLNGDHYKFRVAVNSEGSPDLETETDILAKN 2540
QY 66 ----- 65
Db 2541 PFDRPRGRPEPTDWDSDHVDLKWDPPLSDGGAPIEYQIEKRTKYGRWEPAITVPGGQ 2600
QY 66 ----- 65
Db 2601 TTATVPOLTPNEEYFRVAVNKGSPDPSDASKAVIAKPNLKHIDRALKNLTIKAG 2660
QY 66 ----- 65
Db 2661 QSI SFDVPVSGEPAPVTWHPDNRNRENGRVLNDRNPYOSKLVVKQMERGDSGTETIK 2720
QY 66 ----- 65
Db 2721 AVNANGEDATVKINVIDKPTSPNGPLDVSVDHGDHVTLNWRAPDDGGIPIENYVIEKY 2780
QY 66 ----- 65
Db 2781 DTASGRWVPAAKVAGDKTTAVVDGLIPGHEYKFRVAVNAEGESDPLETFTGLAKOFFD 2840
QY 66 -----AGNIT----- 70
Db 2841 KPGKTNAPEITDWDKXHDVLEWKPPANDGGAPIEYVYVEMKDEPSPFWNDVAHVPAQOTN 2900
QY 71 ----- 70
```

```
Db 2901 ATVGNLKEGSKYEFIRAKNKAAGLGDPPSDASAVAKARNVPPVIDRNSIOEIKVKAGQDF 2960
QY 71 ----- 70
Db 2961 SLNIPVSGEPTPTITWTTFEGTPVESDDRMKLNEDGKTKFHVKRALRSDTGTGTIIKAENE 3020
QY 71 ----- 70
Db 3021 NGDTAEVKVTVLDHPSPSPRGLDVTNIVKDGCDLAWKEPEDDGGAEISHYVIEKODAAAT 3080
QY 71 ----- 70
Db 3081 GRWTACGESKOTNFHVDDLTQGHYKFRVAVNRHGDSDPLEAREAIIAKDPFDRADKPG 3140
QY 71 ----- 70
Db 3141 TPEIVDMDKHADLKWTPPADGGAPIEGVLVEMRTPSGDWVPAVTVGAGELTATVDGLK 3200
QY 71 ----- 70
Db 3201 PGOTYQFRVAKLNKAGESTSPSRPTMVAKPRHLAPKINRDMFVAQVRKAGOTLNFVNV 3260
QY 71 ----- 70
Db 3261 EGEPAPIEWFNLGSPSSGGNTHIDNNTDNNTKLTSTARADSKYKIVATNESGKDE 3320
QY 71 ----- 70
Db 3321 HEVDVNILDIPGAPEGLRHKIDITKESVVLKWDEPLDDCGSPITNYVVEKQEDGGRWVPC 3380
QY 71 -----ANLTS----- 75
Db 3381 GETSDTSLKVNKLSGHEGYKFRVAVNRQOTSAPLTSDBHAI VAKNPFDEPDAPTDTVPD 3440
QY 76 ----- 75
Db 3441 WDKHDVLEWKPPANDGGAPIDAYIVEKDKDFGDWVECARVDGKTKTATADNLTPGETYO 3500
QY 76 ----- 75
Db 3501 FRVAVNKAAGPKSPDPTGNVAVKPRRMAPKLNAGLLDLRIKAGTPIKLDIAFEGEPAP 3560
QY 76 -----SL 77
Db 3561 VAKWKANDATIDTGARADVNTPTSSAIHIFSARVGTGVYKIIIVENEHKGDTAQCNTV 3620
QY 78 LSVCG----- 82
Db 3621 LDVPGTPEGPKIDEIHKEGCTLNWKPPPTDNGGTDVLHYIVEKMDTSRGTWQEVGTFPDC 3680
QY 83 ----- 82
Db 3681 TAKVNLKVPGEYAFRVAVNLOGESKPLEAESEPIIAKQDFVDPDPVDKPEVTDWDKDRI 3740
QY 83 -----WSOTIN----- 88
Db 3741 DIKWNPTANNGAPVTGYIVEKKEGSAIWEAGTKPTGTTFSADNLKPGVEYEFVIAVN 3800
QY 89 -----PEDDTP----- 95
Db 3801 AAGSPDSDPTDPOITKARYLKPILKTASRKIKIKAGFTNLEVDFIGADPTATWTVD 3860
QY 96 ----- 95
Db 3861 SGAALAPELLVDAKSSTTSIFFPSAKRADSGNYKLVKNELGEDEAIFEVIVDRPSAPE 3920
QY 96 ----- 95
Db 3921 GPLEVSQVTKDSCVLNWKPKDDGAEISNYVVEKRDTKTNTWVPVSAFVTGTSITVPKL 3980
QY 96 ----- 95
Db 3981 TEGHEYFRVMAENTFGRSDSLNTDEPVLAKDPFGTGPGRPEIVDTDNHDHIDIKWDPP 4040
```

Qy 96 ----- 95
Db 4041 RDNCGSPVDHYIERDKATGRWIKWNTSPVQGTAFSTRVQKGHTYBYRVVAVNKAQPG 4100
Qy 96 ----- 95
Db 4101 QPSDSSAAATAKPMHEAPKFDLDDCKEPRVKAGEPLVITIPFTASPOQDISWTKGSKP 4160
Qy 96 ----- 95
Db 4161 LAGVETTSQTKLVIPSTRSDSGPVKIKAVNPYGEAEANIKITVIDKPGAPENITYPAV 4220
Qy 96 ----- 95
Db 4221 SRHTCTLNWDAPKDDGGAIEAGYKIEYQEVGSIWIKVPGLISGTAYTVRGLEHQQYRF 4280
Qy 96 ----- 95
Db 4281 RIRAEAVGLSDYCOCPVPVVIKDPDPGAPSTPEITGYDTNOVSLAWNPPRDDGGSFIL 4340
Qy 96 ----- 97
Db 4341 GVVVERFEKGGDWAPVKMPVKCTECIVPGLHENETYQPRVRAVNAAGHGEPSNGSEP 4400
Qy 98 ----- 97
Db 4401 VTCRPYVEKGPADPRVGKITKNSAELTNRLRDGGAPIDGYIVEKKLGDNDWTRCN 4460
Qy 98 ----- 102
Db 4461 DKPVRDTAFENVKLGEKEEYEFVRVIAVNSAGEGSKPSDLVLIBEQGRPIFDINNLIK 4520
Qy 103 ----- 102
Db 4521 ITVRAGETIQIRIPYAGGNPKPIIDLFGNSPIFENERTVVDVNPGEIIVITTTGSKRSDA 4580
Qy 103 ----- 102
Db 4581 GPKISATNKYCKDTCLNVFLDAPKPTGPIRATDIQADAMTLSWRPPKDNGGDAITN 4640
Qy 103 ----- 102
Db 4641 YVVEKRTGGDWVTVGHVPGTTLVRNLDANTPYEFVRRAENQYGVGEPLETDDDAIVAKN 4700
Qy 103 ----- 102
Db 4701 PFDTPGAPQPEAVETSEEAITLQWTRPTSDDGAPIQGVVIEKREVSGSTEMTKAAGNII 4760
Qy 103 ----- 102
Db 4761 DTGHRVTGLTPKTYEFRVAAYNAAOGGEYSVNSVPITADNAPTRPKINMGMLTRDILAY 4820
Qy 103 ----- 102
Db 4821 AGERAKILVPPAASPAPKVFTSKGENKISPTDPRVKVEYSDFLATLTIEKSELTDGGLYF 4880
Qy 103 ----- 102
Db 4881 VELENSQGSDSASIRLVKVDKPPASQHIRVEDIAPDCCTLYMPPSSDGGSPITNYIVEK 4940
Qy 103 ----- YITRFD 108
Db 4941 LDLRHSQGWKEKVSFVRNLNVTGGLIKDNRFRVRAETQYGVSEPCELADVVAKYQ 5000
Qy 109 LELP----- 112
Db 5001 FEVPNQPEAPTVRDKDSTWAELEWDPDRDGGSKIIGYQVQYRDTSSGRWINAKMDLSEQC 5060
Qy 113 ----- 112
Db 5061 HARVTGLRQNGEFEFRIIAKNAAGFSKPSPPSERCOLKSRFGPPGPIHVGAKSIGRNHC 5120

Qy 113 -----DGNROVRG----- 120
Db 5121 TITWMAFLEDGSGKITGYNVEIREYSTLTWTVASDYNVREPEFTVDKLEFNDYFRVVA 5180
Qy 121 ----- 120
Db 5181 INAAKGKIPSLPSGPIKIQESGGSRPOIVVKPBDTAQYNRRAVFTCEAVGRPEPTARWL 5240
Qy 121 ----- 120
Db 5241 RNRGELPSSRYRFEASDGVYKFTTKEVMDIDAGEYTVESNPNYGSDTATANLVVQAPPV 5300
Qy 121 ----- 120
Db 5301 IEKDVNTILPSGLVRLKIYFSGTAPRHSVLNRBEIDMDHPTIRIVEFDDHILITIP 5360
Qy 121 ----- 120
Db 5361 ALSVREAGRYEYTVSNDGSEATTGFMNLNVTGLPEAPOGLHISNIGPSTATLSWRPPTD 5420
Qy 121 ----- 120
Db 5421 GGSKITSYVVEKRDLSKDEWVTTSNVKDMNYIVTGLFENHEYFRVSAQENENGIGAPLV 5480
Qy 121 -----VTOLGG----- 126
Db 5481 SEHPHIIARLPDPPTSPNLNLEIVQGGDYVTLQWQRLPSDGGRLRGYIVEKQEEHDEM 5540
Qy 127 ----- 126
Db 5541 FRCNQNPSPNNYVNPNLIDGRKYRYRFAVNDAGLSDLAELDOTLFOAGSGGEGPKIVS 5600
Qy 127 ----- 126
Db 5601 PLSDLNEEVCRCVTECEISGSPREYRWFPGCKELVDTSKYTLINKGDKQVLIINDLTS 5660
Qy 127 ----- 126
Db 5661 DDADEYTCRATNSSGSTRANLRIKTKPRVFIIPKYHGGYEAOKGETIELKIPYKAYPO 5720
Qy 127 ----- 126
Db 5721 GEARWTKDGEKIENSKFSITDDKATLRIISNAREDEYGEYRVVENSUGSDSGTVNVT 5780
Qy 127 ----- 126
Db 5781 VADVPEPPRPFIIENILDEAVILSWKPPALDGGSLVTNYTIEKREAMGSGSWSPCAKRYT 5840
Qy 127 -----ACSP----- 131
Db 5841 YTTIEGLRAGKQYEFRIIAENKHGQSKPCPTAPVLIPGDERKRRRGYVDVDEQGIVRGK 5900
Qy 132 ----- 131
Db 5901 GTVSSNYDNYVDIWKQYVQPVVEIKHDVLDHVDIHEELGTGAFGVVHRVTERATGNF 5960
Qy 132 ----- 131
Db 5961 AAKFVMTPHESDKETVRKEIQTMSVLRHPTLVNLJHDAPEDDNEMWMIYEFMSGGELFEKV 6020
Qy 132 ----- 131
Db 6021 ADEHNKSEDEAVYMRQVCKGLCHMHENNYVHLDLKPENIMFTTKRSNELKLIDFGLTA 6080
Qy 132 -----WS----- 133
Db 6081 HLDPKQSVKVTGTGTAFAAPEVAEGKPVGYTDMKSVGLSYLLSGLSPFGENDDETL 6140
Qy 134 ----- 133
Db 6141 RNVKSCDNMDDSAFSGISEDGKDFIRKLLADPNTRMTIHOALEHPMLTPGNAPGRDSQ 6200
Qy 134 ----- 133

Db 6201 IPSRVTYKIRDSIKTKYDAMPELPLGLGRISNYSSLKRLKRPQEQYSIRDAFWDSQAQPRF 6260
QY 134 ----- 133
Db 6261 IVKPYGTVEVGOSANFYCRVIASSPPVTVWTKDDRELKOSVKYMKRYNGNDYGLTINRV 6320
QY 134 ----- 133
Db 6321 KGDDKGEYTVRAKNXSGYTKBEIVFLNVTRHSEPLKPEPLEPMKKAPSPRVVEFKERRSA 6380
QY 134 -----CLIT----- 137
Db 6381 PFTFFHLNRLIKNHCKLTCSLQGNPTIEWMKDGHPVDEDRVQVSRSGVCSLEIF 6440
QY 138 -----EDTGF-----DLGVTI 148
Db 6441 NARVDDAGTVTVTATNDLGDV 6462
RESULT 14
S57242
twichin [similarity] - Caenorhabditis elegans
N:Alternate names: myosin-regulating protein
N:Contains: protein kinase (EC 2.7.1.-)
C:Species: Caenorhabditis elegans
C>Date: 28-Oct-1995 #sequence, revision 24-Oct-1997 #text, change 20-Jun-2000
A:Accession: S57242; S07571; S06797; S57218; T27934; T28030
R:Benian, G.M.; L'Hernault, S.W.; Morris, M.E.
submitted to the EMBL Data Library, February 1993
A:Description: Additional sequence complexity within twitching of Caenorhabditis elegans
A:Reference number: S57242
A:Accession: S57242
A:Molecule type: DNA
A:Residues: 1-6839 <BEN1>
A:Cross-references: EMBL:LI0351
A:Experimental source: var. Bristol
R:Benian, G.
submitted to the EMBL Data Library, November 1989
A:Reference number: S07571
A:Accession: S07571
A:Molecule type: DNA
A:Residues: 792-6839 <BEN2>
A:Cross-references: EMBL:X15423; NID:G6897; PIDN:CAA33463.1; PID:G6898
A:Experimental source: var. Bristol
R:Benian, G.M.; Kiff, J.E.; Neckelmann, N.; Moerman, D.G.; Waterston, R.H.
Nature 342, 45-50, 1989
A:Title: Sequence of an unusually large protein implicated in regulation of myosin activity
A:Reference number: S06797; MUID:90044042; PMID:2812002
A:Accession: S06797
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 806-1175; 1178-1998, 'Y', 2000-3040, 'I', 3042-3335, 'I', 3337-5693; 5696-6359, 'I', 6360-6839
A:Cross-references: EMBL:X15423
A:Experimental source: var. Bristol
R:Benian, G.M.; L'Hernault, S.W.; Morris, M.E.
Genetics 134, 1097-1104, 1993
A:Title: Additional sequence complexity in the muscle gene, unc-22, and its encoded protein
A:Reference number: S57218; MUID:93387664; PMID:83971135
A:Accession: S57218
A:Molecule type: DNA
A:Residues: 2-99; 108-194, 'Q', 196-206; 374-468; 658-753 <BEN4>
A:Experimental source: var. Bristol
R:White, S.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z20442
A:Accession: T2934
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 'MGIPGKKCKQ', 19-6839 <WI2>
A:Cross-references: EMBL:Z73897; PIDN:CAA98064.1; GSPDB:GN00022; CESP:2K617.1a
A:Experimental source: clone ZK617
R:Harrie, B.

submitted to the EMBL Data Library, May 1996
A:Reference number: Z20458
A:Accession: T28030
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 'MGIPGKKCKQ', 19-6839 <WI2>
A:Cross-references: EMBL:Z73899; PIDN:CAA98081.1; GSPDB:GN00022; CESP:2K617.1a
A:Experimental source: clone ZK829
C:Comment: Lack of unc-22 leads to a constant twitching of the body muscles.
C:Genetics:
A:Gene: unc-22; CESP:2K617.1a
A:Map position: 4
A:Introns: 18/3; 69/3; 143/2; 176/3; 264/2; 387/3; 413/2; 471/1; 516/3; 550/3; 582/3;
152/3; 6691/3; 6776/1; 6808/3
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology
C:Keywords: ATP; autophosphorylation; duplication; muscle; phosphotransferase; serine
F1806-898, 899-990, 991-1083, 1084-1175, 1178-1273, 1474-1567, 1770-1864, 2066-2158, 2358-241
96-5790, 6263-6356, 6386-6478, 6541-6635, 6649-6742, 6745-6838/Region: motif 2
F1274-1372, 1373-1473, 1568-1670, 1671-1769, 1865-1964, 1965-2065, 2159-2258, 2259-2357, 241
23, 4215-4313, 4314-4415, 4416-4516, 4612-4710, 4711-4811, 4908-5009, 5010-5109, 5110-5210, 51
F:5940-6197/Domain: protein kinase homology <KIN>
F:5948-5956/Region: protein kinase ATP-binding motif
F:5971/Active site: Lys Metastatus predicted
Query Match 28.6%; Score 221; DB 2; Length 6839;
Best Local Similarity 1.3%; Pred. NO. 1.7e+02;
Matches 81; Conservative 28; Mismatches 32; Indels 6181; Gaps 25;
QY 3 GGILHL----- 8
Db 154 GGLYHAIFSDLDGDTYLCQLEIRGSSDAGYRCNIRNDOGETNANLANFEEPDPSE 213
QY 9 ----- 8
Db 214 QERKRSTASPRSSRGPRSPSPKSMKSRCTPKTLKPREGSPSKLRSTSTPVNE 273
QY 9 ----- 8
Db 274 EYVSQSESRSSRTDKMEVDQVSGASKRPDGLPPPGGDEKLRAGSPSTRKSPRSKASP 333
QY 9 ----- 8
Db 334 TPKRGSSAGGAAGTTGASATSATSGSASSASADSKYTRPPVLEASRSQTRIGG 393
QY 9 ----- 8
Db 394 SVVLEVOMQCHSSTIEWYRDGTLVRNSSEYSQSFNSTAKLVNKLTEKSGLYKCHAK 453
QY 9 ----- 8
Db 454 CDYGEQSSAMVKIEQSDVEELMKHKDAEDYQKEQKSTLQAEKKRVARRSKSKS 513
QY 9 -----ELL 11
Db 514 KSPAOKKSTTSSEGRQEAHEVHKRSSVRPDPDEESQDLEIPSSGLTIPERRRELL 573
QY 12 VAVG----- 15
Db 574 GQVGESEDSVESISELPSFAGKPRKTDKPKKVSIAPVSTNKSDDDFSTPRRSSI 633
QY 16 ----- 15
Db 634 DMRRESVQETLEKSTPLVPSGASGAPKIVEPVENVVVENETAILTKCVSGSPAPTF 693
QY 16 ----- 15
Db 694 WFKGSREVISGGRFKHITDGKHTVALALLKCRSQDEGPTYLTIENTVHGTDSDVKLLVT 753
QY 16 ----- 15
Db 754 SDNGLDFRAMLKXHSOAGFOKDGEGGAGGGGKPKMPTEARROSLFPCKKVEKWDP 813
QY 16 -----PD 17


```
QY 71 ----- 70
Db 3029 NGTDTAEVKTVLDFHPSRGLDVNTIVKDGCDLAWKEPEDDGGAEISHYVIEKQDAAT 3088
QY 71 ----- 70
Db 3089 GRWTAGESKDTNFHVDLTLQGHEYKFRVAVNRHGDSDPLEAREAIIAKDPFDRADKPG 3148
QY 71 ----- 70
Db 3149 TPEIVDMXKDHADLKWTPPADGGAPLEGYLVEMRTPSGDWVPAVTVGAGELTATVDGLK 3208
QY 71 ----- 70
Db 3209 PGQTYOFRVAKLNKAGESTPSDPRMTVAKPRHLAKINRDMFVAORVAKAGQTLNFDVNV 3268
QY 71 ----- 70
Db 3269 EGPAPKIEWFLNGSPSSGGNTHIDNNTDNNTKLTTKSTARADSGYKIVATNESGKDE 3328
QY 71 ----- 70
Db 3329 HEVDVNILDIPGAPEGLRHKOITKESVVLKWDPELDDGSPITNYVVEKQEDGGRWVPC 3388
QY 71 ----- 75
Db 3389 GETSDTSLKVNKLSEGHEYKFRVAVNRQGTSAPLTSDHAIVAKNPFDEPDAPTDTVPD 3448
QY 76 ----- 75
Db 3449 WDKDHVLEWKPPANDGGAPIDAYIVEKKDKFGDWECARVDGKTTKATADNLTPGETYQ 3508
QY 76 ----- 75
Db 3509 FRVAVNKAGPGKSDPTGNVAVAKPRRMAPKLNLAGLLDLRIKAGTPIKLDIAFEGEPAP 3568
QY 76 ----- 82
Db 3569 VAKWKANDATIDTGARADVNTNTSSAIHIFSARVGDGTGVYKIIIVENEHCKOTAQCNVTV 3628
QY 78 LSVCG----- 82
Db 3629 LDVPGTPEGKIDEIHKEGCTLNWKPPTDNGGTDVLHYIVEKMDTSRGTVQEVGTGPD 3688
QY 83 ----- 82
Db 3689 TAKVNKLVEGKEYAFRVAVKAVNLOGESKPLEAEPIIAKQFVDPDPVKPEVTDMDKRI 3748
QY 83 ----- 88
Db 3749 DIKWNPTANNGGAPVTGYIVEKKEGSAIWTEAGKTPGTTFSADNLKPGVEYEFRIAVN 3808
QY 89 ----- 95
Db 3809 AAGPSDPSOPTDQITKARVLKPKILTASRKIKIKAGFTHNLEVDFIGAPDPTATWTVDG 3868
QY 96 ----- 95
Db 3869 SGAALAPELLVDAKSSTSIFFPSAKRADSGNYKLVKNELGEDEAIFEIVQDRPSAPE 3928
QY 96 ----- 95
Db 3929 GPLEVSDVTKDSCVLNWKPPKDDGGAEISNYVVEKRDTKTNWVPSAFVTGTSITVPKL 3988
QY 96 ----- 95
Db 3989 TEGHEYEFRVMAENTFGRSDSLNTEPVLAKDPFGTGPGRPEIIVDTDNHDHIDIKWDDP 4048
QY 96 ----- 95
Db 4049 RDNGGSPVDHYDIERKDAKTGRWIKVNTSPVQGTAFSDTRVQKHTYEVYRVAWNKAGPG 4108
```

```
QY 96 ----- 95
Db 4109 QPSDSSAAATAKPMHEAPKFDLDLGKEFRVAKAGPLVITIPFTASPOQDISWTEGGKP 4168
QY 96 ----- 95
Db 4169 LAGVETTSQTKLVIPSTRRSDDSPVKIKAVNPYGEAEANIKITVIDKGPAPENITYPAV 4228
QY 96 ----- 95
Db 4229 SRHTCTLNWDAPKDDGGAEIAGYKIEYQEVGSOIWDKVPGLISGTAYTVRGLEHGOQYRF 4288
QY 96 ----- 95
Db 4289 RIRAEVAVGLSDYCOGVPVWIKDPFPPGAPSTPEITGYDTNOVSLAMNPPRDGGSPIL 4348
QY 96 ----- 97
Db 4349 GYVVERFEKRGGDWAPVXMPVMWKGTECIVPGLHENETVQFRVAVNAAGHGEPSNGSEP 4408
QY 98 ----- 97
Db 4409 VTCTPYVEKPGADAPRVCKITKNSAELTNWRLRDGGAPIDGYIVEKKKLGDNWTRCN 4468
QY 98 ----- 102
Db 4469 DKPVRDTAFEVKNLGEKEBEYFRVAVNSAGEPSKPSDLVLIBEQGPRPIFDINNLKD 4528
QY 103 ----- 102
Db 4529 ITVRAGETIQIRIPYAGNPKPIIDLFNNGSPIFENERTVVDVNPGEIVITTTGSKRSDA 4588
QY 103 ----- 102
Db 4589 GPKISATNKYCKDTCKLVFVLDAPKPTGPIRATDIOADAMTLSWRPCKONGGDAITN 4648
QY 103 ----- 102
Db 4649 YVVEKRTPGDWMVTGVHPVGTTLRVRLDANTPYEFRVRAENQYGVGEPLETDDAIVAKN 4708
QY 103 ----- 102
Db 4709 PFDTPGAPQPEAVETSEBAITLWTRPTSDGAPIQGYVIEKREVSTEWTKAAGFNIL 4768
QY 103 ----- 102
Db 4769 DTGHRVTGLTPKKTVEFRVAAYNAAGQGEYSVNSVPIADNAPTRPKINMGMLTRDILAY 4828
QY 103 ----- 102
Db 4829 AGERAKILVPPAASPAPKVTFSKGENKISPTDPRVKVEYSDFLATLTIEKSELTDGGLYF 4888
QY 103 ----- 102
Db 4889 VELENSQGSDSASIRLKVVDPKASPQHIRVEDIAPDCCTLYWMPSPSSDGGSPITNYIVEK 4948
QY 103 ----- 112
Db 4949 LDLRSDGKWEKVSFVRNLNTVVGGLIKDNRYRFRVRAETOYGVSEPCELADVVAKYQ 5008
QY 109 LELP----- 112
Db 5009 FEVPNOPEAPTVRDKDSTWAELEWDPDRDGGSKIIGYQVQYRDTSSGRWINAKMDLSEOC 5068
QY 113 ----- 112
Db 5069 HARVTGLRONGEFEPRIIAKNAAGFKSPSPSERCOLKSRFGPPGPIIHVGAKSIGRHNH 5128
QY 113 ----- 120
Db 5129 TITWMAPLEDGGSKITGYNVIREYVSTLWTVASDVNVRPEFTVDKLRFNDEYPRVVA 5188
QY 121 ----- 120
```


Db 5189 INAGKIPSLPSGPIKIQESGSRPQIVVKPEDTAQPNRRVFTCEAVGRPEPTARWL 5248
Qy 121 ----- 120
Db 5249 RRGELPSSRYRFEASGVYKFTIKEVMDIDAGEYTVESNPGSDTATANLVQAPPV 5308
Qy 121 ----- 120
Db 5309 IEKVPNTILPSGDLVRLKIYFSGTAPPRHSLVLRNREIDMDHPTIRIVEFDDHILITIP 5368
Qy 121 ----- 120
Db 5369 ALSVREAGRYEYTVNSDGEATTGFWLNVTCGLPEAQGPLHISNIGPSTATLSMRPPVTD 5428
Qy 121 ----- 120
Db 5429 GSKITSYVEKRLDSKDEWTVTSNVKDMNYIVTGLFENHEYFRVSAQNGIGAPLV 5488
Qy 121 ----- 126
Db 5489 SEHPILARPDPPTSPUNLEIVQGGDYVTLWSORPLSDGGGRGLRGYIVVEKQEEHDEW 5548
Qy 127 ----- 126
Db 5549 FRCNQPPNNYVNPVLIDGRKYRYRVFAVNDAGLSLAELDQTLFQASGSGGPKIVS 5608
Qy 127 ----- 126
Db 5609 PLSLDNEVGRCVTFECISGSPREYRWFKCKELVDTSKYTLINKGDKQVLIINDLTS 5668
Qy 127 ----- 126
Db 5669 DDADEYTCRATNSSGTRSTRANLRKTKPRVFIPPKYHGVEAOKGETIELKIPYKAPQ 5728
Qy 127 ----- 126
Db 5729 GEARWKDGEKIENKSFIITDDKFATLRISASREDYGEYRVVENSVDGSGTVNVT 5788
Qy 127 ----- 126
Db 5789 VADVPEPRFPIENILDEAVILSWKPPALDGGSLVTNTYTIKREAMGSGSPCAKRYT 5848
Qy 127 ----- 131
Db 5849 YTTIEGLRAGQYEFRIIAENKHGOSKCEPTAPVLIPGDERKRRRGYVDVDEQGIKVRGK 5908
Qy 132 ----- 131
Db 5909 GTVSSNYDNYVDIWKQYYPQVVEIKHDHVDLHDYDIHEELGTGAGVVRVTERATGNVF 5968
Qy 132 ----- 131
Db 5969 AAKFVMTPHSEDKETVRKEIQTMSVLRHPTLVNLHDAPEDDQNMVMYEFMSGGELFEK 6028
Qy 132 ----- 131
Db 6029 ADEHNKMSDEAVEYMQVCKGLCHMHENNYVHLDPENIMFTTKRSNELKLIIDFGLTA 6088
Qy 132 ----- 133
Db 6089 HLDPKQSVKVTGTAEFAAPEVAEKGKPVGYTDMWSVGLSYLLSGLSPFGGNDDETL 6148
Qy 134 ----- 133
Db 6149 RNVKSCDNMDDSAFGISGDGKDFIRKLLADPNTRMTIHQALEHPWLTPGNAPGRDSQ 6208
Qy 134 ----- 133
Db 6209 IPSSRYTKIRDSIKTKYDAPPEPLPGLGRISNYSSLRKHPRQEYSIRDAFWRSEAPRF 6268
Qy 134 ----- 133

Db 6269 IVKPYGTGEGOSANFYCRVIASSPPVVTWTHKDDRELKQSVKMKRYNGNDYGLTINRV 6328
Qy 134 ----- 133
Db 6329 KGDDKGEYTVRAKNSVGTKEEIVFLNVTRHSEPLKPEPLEPMKKAPSPRVEEFKERRSA 6388
Qy 134 ----- 137
Db 6389 PPTFFHLNRLNIQKNHOCKTSLQGNPNPTIEWMKDGHVDEDRVQVSPRSGVCSLEIF 6448
Qy 138 ----- 148
Db 6449 NARVDDAGTYVTATNDLGDV 6470
RESULT 15
T30283
polyketide synthase - Streptomyces sp. (strain MA6548)
C:Species: Streptomyces sp.
A:Variety: strain MA6548
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C:Accession: T30283
R:Motamedi, H.; Cai, S.J.; Shafiee, S.J.; Elliston, K.O.
Eur. J. Biochem. 244, 74-80, 1997
A:Title: Structural organization of a multifunctional polyketide synthase involved in
A:Reference number: Z20806; MUID:97217427; PMID:9063448
A:Accession: T30283
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6420 <MOT>
A:Cross-references: EMBL:Y10438; NID:e1014806; PID:e290681; PIDN:CAA71463.1
C:Genetics:
A:Note: fkbA
C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier pr
C:Keywords: carrier protein
F:51-433/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <AS1>
F:1930-2325/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <AS2>
F:3391-3462/Domain: acyl carrier protein homology <ACP1>
F:3505-3900/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <AS3>
F:3983-4234/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
F:5307-5378/Domain: acyl carrier protein homology <ACP2>
F:5431-5831/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <AS4>
F:5932-6206/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
F:6293-6364/Domain: acyl carrier protein homology <ACP3>
Query Match 28.5%; Score 220; DB 2; Length 6420;
Best Local Similarity 1.5%; Pred. No. 1.5e+02;
Matches 88; Conservative 25; Mismatches 35; Indels 5843; Gaps 27;
Qy 1 AAGGI-----LHLE----- 9
Db 360 AAGVAGIIMVLAHRHGLTLPPTLHVDTSPSHVDWTTGDVELLTQARPWQTNPRHAGI 419
Qy 10 -----LLVAV----- 14
Db 420 SFEVSGTNAHVIESHPRTAPDTGSGSTHPVPLLSARTPEALDEHTTIRAFLDNP 479
Qy 15 -----GPDV----- 18
Db 480 GADRVTVQTLARRTPPEHRAVLGLDTLITADPNAGSGPVVYVSGSTLHPHTGHOLAA 539
Qy 19 ----- 18
Db 540 TYSVFADWAGEVLGHLNADOGPATHFAHQIALTALLRSWGITPHAVIGHSLGEISAACAA 599
Qy 19 ----- 18
Db 600 GVLSIGDASALLAARSRLMDELPTGGAMVTVLTSEENALRALRPGVEIAAVNGPHSVLS 659
Qy 19 ----- 18
Db 660 GDEGVLDAVQQLGIHRLPTRHAGHSARMDPLVAPLLEASGLTYHOPHTAIPEDPTTA 719

QY 19 -----FOAHQE----- 24
Db 720 AYWAROVROVRFQAAERYPGATFLEIGNQDLSPVVDGIPQTGTGPEVQALHTALAR 779
QY 25 ----- 24
Db 780 LHTRGGVMDPTVLGSDRAPVALPTYPFOHKDYWLRTAQVDVTGAGQEKVAHPLLGAUV 839
QY 25 ----- 24
Db 840 ALPGTGTVLTRVSLASHPNLGGHAVNGTVLLPGAFLAVRAGDETGSDLLHELVE 899
QY 25 -----DTER----- 28
Db 900 TPLTLPATGSAISVEAEPPDTGCRRAVTVHARADGAALWTRHATGFLGKAATAATTAE 959
QY 29 ----- 28
Db 960 PAPMPAGARQVDVYERFAGIGYAGSSFOGLLAWRDGTVHAVALPDEQVPDAA 1019
QY 29 ----- 28
Db 1020 RFTLHPALLDAFQAGALSALDAPGEAARLPFSFQDVRTHAAGTRARVTVSRDGRSTV 1079
QY 29 ----- 28
Db 1080 RMTGEDQPAEVGSVSRPYAGSGSDDELLRPVMTLPMAPSAADPRVEVLDADPLD 1139
QY 29 ----- 28
Db 1140 GEPAAATRELTARVLDLSQHLRSATDDVTLVVRIGTGAAAAATAGLVRSAQENPGRFVL 1199
QY 29 ----- 28
Db 1200 VEAPDTPADLLTACAALDEPQLAVRDGVLFPAPLVRMSDPVHGPLPLPDGDLWLLAHSAS 1259
QY 29 ----- 28
Db 1260 GVLQDVTLAADGTPRRALEAGEVRVDVRAAGLNFRDVLIALGTYDGATALLGGEAGVVE 1319
QY 29 -----YVLT 32
Db 1320 TGPGVADLAPGDRVFGLTRGGIGPTAVTDRRWLARIPOGWTFFATAASVPIAFATAWYGLV 1379
QY 33 NL-----NIGAE----- 40
Db 1380 DLAGLRAGEKVLVHAATGGVMAATQVAREHLGAEVYATAGTGKQHVLRAGLPGTHVADS 1439
QY 41 -----LRDP----- 44
Db 1440 RTTAFRTDPPRMDVNLSTGBEFIDASLELLDADGRFVEMGRAELRDPATIAPAYLPFDL 1499
QY 45 -----SLGAQFR----- 51
Db 1500 LDAGTDRIGEILGELLRLFDAGALTPLPVRAWDIROARDALGYMSRARHIGKNVLTLP 1559
QY 52 -----VHL----- 54
Db 1560 LDPEGAIVITGSGTLAGILARHVSQHVYLLSRSPPEGTGPIHLPCDVSRRQQLAAAL 1619
QY 55 ----- 54
Db 1620 RRIDRPITAVVHTAGALDDGTLPSTLPGRFDTLPRKADGAWHLHELTKDQDLAAFLVYS 1679
QY 55 ----- 54
Db 1680 SAAGVLNAGOGNYAAANAFDALAEORRAAGLPALSIAWGLWEDASGLTAELSEADQR 1739
QY 55 ----- 54
Db 1740 MRRSGFRITARQGMRLYEARRTGSVPVVAALDDAPDVPLLRGLURRTTVRRAAVRERS 1799
QY 55 ----- 54

Db 1800 LADSPCCPTTSAPTSPSRSSWNSTATVLGHLGAEDIPATTTFKELGIDSLTAVOLRNAL 1859
QY 55 ----- 54
Db 1860 TTATGVRLNATAVDFPTTPRALAARLGDELACTRAPVAARTAAATAAHDEPLAIVGMACR 1919
QY 55 ----- 54
Db 1920 LPGVASPOELWRLVASGTDAITFPPADRGWDVLDALYDPPDAIGKTFVRHGGFLDGTG 1979
QY 55 ----- 54
Db 1980 FDAFFGISPREALAMPQORVLLTSEAFESAGITPDAARSGDTGVFIGAFSYGYGTG 2039
QY 55 ----- 54
Db 2040 ADTNFGATGTSQTSVLSGRLSYFYGLEGPSVTVDTACSSSLVALHQAGOSLSRSGCSLAL 2099
QY 55 ----- 54
Db 2100 VGGVTVMASPGGFVBFSRQGLAPDGRAKAFGAGADGTSFAEGAGALVVERLSDAERHGH 2159
QY 55 ----- 54
Db 2160 TVLALVRGSAANSDBGASNGLSAPNGPSQSERVTHQALANAKLTPADVDVAVEAHGTGTRLGD 2219
QY 55 -----VKMV----- 58
Db 2220 PIEAQLLATYQDQTRATPLLLGSLKSNIGHAASGVAIGIIMVQAIRHGELPPTLHADE 2279
QY 59 ----- 58
Db 2280 PSHVDWTAGAVELLTSARPWGTGRPRRAAVSSFCVSGCTNAHIIIEAGPVKTPVEAGA 2339
QY 59 ----- 58
Db 2340 IEAGVEVGPVEAGPLPAAPSAGEDPLLVARSPEALDEQIGRLRAYLDTGPGVDRA 2399
QY 59 ----- 58
Db 2400 AVAQTLLARRTHTRAVLLGDTVIGAPPADQADELVFVYSGGTQHPANGEOALAAFPVF 2459
QY 59 ----- 58
Db 2460 ADAMHDALRLDDPDPHDPTRSQHTLFAHQAAFTALLRSWDITPHAVIGHSLGEITAAVA 2519
QY 59 -----ILTE----- 62
Db 2520 AGILSLDDACTLITTRARLMHTLPPPGAMVTULTSEEARQALRPGVEIAWFGPHSVVL 2579
QY 63 ----- 62
Db 2580 SGDEDAVLDOAORLGIHRLHRLPAPHAGSAHMEPVAAELLATTRELYRDPHTAIPNDPTT 2639
QY 63 ----- 62
Db 2640 AEYWAQVNRNPVLFRAHTQRYPDVFEVIGPGODLSPLVDGIALQNGTADVEHALHTALA 2699
QY 63 ----- 62
Db 2700 RUFTRGATLDWSRIILGGASRHPDVPVSVAQRPPYWIESAPPATADSGHPVLGTGVAVAG 2759
QY 63 ----- 62
Db 2760 SPGRVFTGVPAGADRAVPIAELALAAAADATCATVEQLDVTSPGSGSARGRATAQTWVD 2819
QY 63 ----- 62
Db 2820 EPAADGRRRFTVTRVGDAPWTLHAEGVLRPGRVPOPEAVDTAWPPPGAVPADGLPCAHR 2879
QY 63 ----- 62

```
Db 2880 RADQVFEAEVSDPDGFVAHPDLLDAVFSVAGDGSRQQTGWRDLAVHASDATVLRACLTR 2939
Qy 63 ----- 62
Db 2940 RDSGVVLAAPFDGACMPVLTAESVTLGEVASAGSDSDGLRLLEWLPVAEAYDGADEL 2999
Qy 63 PEG----- 65
Db 3000 PEGYTLITATHPDDPDPTNPHTPTRHTQTTRVLTAQHLLITTNHTLIVHITTDPPG 3059
Qy 66 ----- 65
Db 3060 AAVTGLTRTAQNEHPCRILHETHHPTPLPLTQLTTLHQPHLRTNNTLHTPHLTPIIT 3119
Qy 66 ----- 65
Db 3120 HHNTTTTTPNPNHAILITGSGTLAGILARHLNHPHTYLLSRTPPPPTPGTHIP 3179
Qy 66 ----- 65
Db 3180 COLDPTQITQALTHI PQPLTGI FHTAATLDDATLNTLPQHLTTTLQPKADAAWHLHHH 3239
Qy 66 ----- 65
Db 3240 TQNOPLTHFVLYSSAAATLSPGOANYAANAFLDALATHRHTQGPATTIAGMWHTTT 3299
Qy 66 ----- 65
Db 3300 TLTSOLTSDDRIRRRGGFLPISDDEGMRLYDAAVSGEDEFVLAAMDPAQPMAGDVVPI 3359
Qy 66 ----- 65
Db 3360 LSGLRKSARTARTQOTFAQRALAEPLAADRDTALVTVUSDATAAVLGHADASGIAPTTTF 3419
Qy 66 ----- 65
Db 3420 KDLGIDSUTAELNRNLAETGLRLSATMVFDHPTPRVLAAELRTDLFGTAAPLPARTAR 3479
Qy 66 ----- 65
Db 3480 TRHDEPLAIVGMACRLPGGVTS PEDMLRLVAGSTDAITEFPTDRGWDIDRMFDPDPDAPG 3539
Qy 66 ----- 65
Db 3540 KTYVRHGGFLSEAAAGFDAAPFGISPREAWANDPOQRVILETVWEAFENAGIVPTLRGSD 3599
Qy 66 ----- 65
Db 3600 TGVFMGAFSHGYGAGDDLGGFGATATQNSVLSGRLSYFFGMEGPAVTIDTACSSSMVALH 3659
Qy 66 ----- 65
Db 3660 QAAQSLRDGCSLALAGGVTYMPTPLGYVEFCRQGLAPDGRAKAFAGADGTSFSEGAG 3719
Qy 66 ----- 68
Db 3720 VLVVERLSDAERNGHTVLALVRSSAVNQDGASISAPNGSQORVIRQALDKAGLAPAD 3779
Qy 69 ----- 68
Db 3780 VDVVEAHGTGTPLDPIEAQAI IATYGQDRTDTPLYLGSVKSNIIGHTQTOTAGLAGVIKWMV 3839
Qy 69 ----- 68
Db 3840 AMRHGLLPKTLHVDEPSSHVDWSAGAVELLTEARWPDPDRPRAGVSSLSIGSTNAHVI 3899
Qy 69 ----- 68
Db 3900 LEGVAESSVRVSESSGLVPLPVARSATESSLALQVERLGEYVRGARDLAAVADGLVRGTV 3959
Qy 69 ----- 68
Db 3960 FGHRVAVLLGESTVAGVAGAESRTVFVFPQGQSQVMGMGRELMSGCEVFAARMRECAAVLE 4019
```

```
Qy 69 ----- 68
Db 4020 PYTGMDDLVDLGEAVVAERVELQPAWAVASVLAALWQAHGVSPDAVIGHOSGEIAAAC 4079
Qy 69 ----- 68
Db 4080 VAGALSLEDAARIVALRSQTIAAHLAAGRGAMASIALPATAVETVEGVVVAARNGPSTVV 4139
Qy 69 ----- ITANLTSSLLSVCCWS 84
Db 4140 AGDPSAVERVLARYEAEGRVRRRIADVASHTPHVEAIQOLADVLGDITSSAPSVPMWS 4199
Qy 85 ----- 84
Db 4200 TVDGGWTEPAGDDYWRNLQRPVMDTAIGELDGSFIECSAHPVLLPALDOERTVASL 4259
Qy 85 ----- 84
Db 4260 RTDDGGHDFELTALAQAQWTOGADVDMTTLIAPAPDRLLDLPTYFFDHKRYWIEATGAADL 4319
Qy 85 ----- 84
Db 4320 TALGLTDTAHPMLAATAALPTNDGTVLTRISLRTHPWLADHTIODTVLLPGTAFLELVM 4379
Qy 85 ----- QTIN----- 88
Db 4380 RAGDEVCDDTIDELVETPLTLPVTGAVDLTVTVEQADATGHRPVSVHARPAGTDTWTRH 4439
Qy 89 ----- 88
Db 4440 ATGTLGRTEDTAPDTNSFPQWPTGPTALDVTEFYAQLAALGYRFVPTFRGLRAAMRAGD 4499
Qy 89 ----- PED----- 91
Db 4500 TVYAEVALPEDRAADADHFGVHPALLDAALQSGSLLMLESDEGQVQLPFSWHGVRFHAT 4559
Qy 92 ----- 91
Db 4560 GATSLRVAIVPGPDGURLRAADSENHPVATIDALVTRPPADLAPSDPVLRVGWTAVPVPA 4619
Qy 92 ----- 91
Db 4620 EVRPSDADVLMHSDDADPIQETRDLTTRVLDALLTTEGTLIVQVTTGGLAAMAAAGLVRT 4679
Qy 92 ----- DTDPG----- 96
Db 4680 AOSEQPGRFFLVETDPGEVLDGARCDMIAALGEPRLRLRGRNFEAARLVRAAAAALTTPDPT 4739
Qy 97 ----- 96
Db 4740 GSWQLRQSATGSLDDVAIVPTGEPDRPLGAGEVRIAVRAAGLNFRDRTVALGVVADRPL 4799
Qy 97 ----- 96
Db 4800 GSEAAGVWLETGPGVHDLAPGDRVLGMLAGAFGPVAITDORLLGRMPDGTWTFPQASVVT 4859
Qy 97 ----- HADLVLYTR----- 106
Db 4860 VFATTWGLVDLAELRPGKVLIIHAAATGVGAAAQVQIARHLADADVVATATSTPKOHLDDLE 4919
Qy 107 ----- F 107
Db 4920 PTHIADSRRTAFADTFPPADVVVLSLTGFLDASAQLLTPGGRFIEMGKTDIRHHVQOPF 4979
Qy 108 DLEL----- 111
Db 4980 DLMVAGPDRLQIIITOLLDLFANGTLHLPLVPHAWDVQRQAREAFGMWSSGRHTCKLVLTVP 5039
Qy 112 ----- PDG----- 114
Db 5040 RPLDPEGAVIITGGSGTLAGHLARHLDPHTYLLSRTPPPPTTTPGTHLPCDVSDPTQLAT 5099
```

```

QY 115 ----- 114
Db 5100 ALAQIPOP LTGIFHTAATLDDALLHNLTPQIDITVLOPKADAAMHLHHHTHATDLAAAF 5159
QY 115 ----- 120
Db 5160 VLYSAVAGMSPGOGNYAANAFUDALAEHQAMGLPAQSLANGMWADVSAFTAKLTDA 5219
QY 121 ----- 126
Db 5220 DRGRIIRGGFPPLSAACMRLFDISAARTSRSSRRPSDVTRLDGPVAPLLRGLVAHRSG 5279
QY 127 ----- 126
Db 5280 PVRAVNADDEPLAARLACHTAAREQRMQEVVLRQAAVLAAYGLGEQVAADRPRDLG 5339
QY 127 ----- 126
Db 5340 FDSLTAVDLRNLAAETGLRLPTTVFVSHPTAEALATHLELLDAPALTTGAPLPVTA 5399
QY 127 ----- 126
Db 5400 APGTAARDQDEPIAIVAMACRLPGVTSPEELWRLVESGTDAITTAPDRGWDLDALYDP 5459
QY 127 ----- 126
Db 5460 DPDAVGKAYSIRGGFLEGAEEFAAFDISPRESIGMDPQORLLLETAEIAERGRINPA 5519
QY 127 ----- 129
Db 5520 SLHGREIGVYVGAAGGYGLCAEDTEGNAITGGSTSLSGRLAYVGLGPAVTVDTACS 5579
QY 130 ----- 129
Db 5580 SSLVALHLAGOGLRGECELALAGVSVLSSPAFVFSRQRLAAGDRCKSPGSGADGT 5639
QY 130 ----- 129
Db 5640 TWAEQVGLVLEQLSDAERLGHVLAUVGSAVTSDCASNGLTAPNGLAQORVIRKALAA 5699
QY 130 ----- 132
Db 5700 AGLTAADVLDVEGHGTGTRLDGPVEADALLATYQNGRGPVWLGSUKSNIGHATAAGVA 5759
QY 133 ----- 132
Db 5760 GVIKTVOAIGAGTMPRTLHADESPAVDWTAGVSLTGNRPWMTSVPGERSPRSGSA 5819
QY 133 ----- 135
Db 5820 GRTLTSLSNSTVSPWRRDLARNPCRLGCCPRPLPCGRRPPVCAIIRPSGPDPLD 5879
QY 136 ----- 135
Db 5880 IGYAQTSAQFTHRAAEAAASPDGFRAGLDSVADGAEPSSAEPPRSARRELPDQOGA 5939
QY 136 ----- 135
Db 5940 QRVCMGRELHGRFPVFAAAWDEVSADAFGKLEHSPDVFHGEHGLAHDLYAQVGLFTL 5999
QY 136 ----- 135
Db 6000 EVALLRLEHMGVRPDVVVGHSGVEVTAAYAAGVLTADATTLIVARGRALRPPGAMT 6059
QY 136 ----- 135
Db 6060 AVEGSPAECVCAFTDLDAIAVNGSPVAVLTGAPDDVAAPEREWAAAGRRAKRLDVGHAFHS 6119
QY 136 ----- 135
Db 6120 RHDVGALDDFRVGLSFAAGARLPVVSTTTTGRDAAGDLATPEHWLRHARRPVLYADAVR 6179
QY 136 ----- 135

```

```

Db 6180 ELADLGVMNFVAVGPGSALASAASENTGGSAGTYHAVLRARTGEESAALTAVAEHLHAHCA 6239
QY 136 ----- 135
Db 6240 PVDLAAVLAGGRPVDLPVYPFOHRSYWLAPAVGGSPATAVPDTGRPTPDPPDPLTVAEIV 6299
QY 136 ----- 148
Db 6300 RRRAAALLGIADPGVDADTTFFALGFDSLAVORLNRNOLTAATGLDLPTAV 6350

```

Search completed: March 20, 2003, 12:41:52
Job time : 204 secs

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	237	30.7	997	1	ATS7_HUMAN	Q9UKP4 homo sapien
2	239	29.6	1593	1	AT12_HUMAN	P58397 homo sapien
3	226	29.2	5065	1	EPPL_HUMAN	P58107 homo sapien
4	224	29.0	5255	1	BACA_BACLI	O68006 b bacitraci
5	222	28.7	5217	1	HTS1_COCCA	Q01886 cochlicobolu
6	216	27.9	4829	1	BIR6_HUMAN	Q9NR09 homo sapien
7	215	27.8	4349	1	FAT2_HUMAN	Q9RVQ8 homo sapien
8	212	27.4	5035	1	RYR1_PIG	P16960 sus scrofa
9	211	27.3	4351	1	FAT2_RAT	O88277 rattus norv
10	210	27.2	5038	1	RYR1_HUMAN	P21817 homo sapien
11	210	27.2	6486	1	TYCC_BACBR	O30409 b tyrocidia
12	209	27.0	3649	1	ACVS_NOCLA	P27743 nocardia la
13	209	27.0	5147	1	FAT_DROME	P33450 drosophila
14	208	26.9	4303	1	PKD1_HUMAN	P98161 homo sapien
15	207	26.8	4590	1	FATH_HUMAN	Q14517 homo sapien
16	206	26.6	4969	1	RYR2_RABIT	P30957 oryctolagus
17	205	26.5	4639	1	DYHC_DROME	P37276 drosophila
18	204	26.4	890	1	RYR8_HUMAN	Q9UP79 homo sapien
19	204	26.4	5037	1	ATR1_RABIT	P11716 oryctolagus
20	203	26.3	4967	1	RYR2_HUMAN	Q92736 homo sapien
21	203	26.3	6359	1	BACC_BACLI	O68008 b bacitraci
22	202	26.1	1629	1	ATS9_HUMAN	Q9P2N4 homo sapien
23	202	26.1	3988	1	POLG_EVDVN	P19711 bovine vira
24	200	25.9	4870	1	RYR3_HUMAN	Q04561 lelystad vi
25	200	25.9	4870	1	RPOA_DELV	Q15413 homo sapien
26	199	25.7	967	1	ATS1_RAT	Q9WUQ1 rattus norv
27	199	25.7	3898	1	POLG_HCVB	P21530 hog cholera
28	198	25.6	3396	1	PGCV_HUMAN	P13611 homo sapien
29	198	25.6	4128	1	PRKD_HUMAN	P78527 homo sapien
30	198	25.6	4540	1	DYHC_PARTIE	Q217171 paramecium
31	198	25.6	5327	1	ACF7_MOUSE	Q9GXZ0 mus musculu
32	197	25.5	968	1	ATS1_MOUSE	P97857 mus musculu
33	196	25.4	860	1	ATS6_HUMAN	Q9UKP5 homo sapien

```

DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSP1; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 232 BY SIMILARITY.
FT CHAIN 233 997 ADAMTS-7.
FT SITE 204 204 CYSTEINE SWITCH (POTENTIAL).
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 389 389 BY SIMILARITY.
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 462 537 *DISINTEGRIN-LIKE.
FT DOMAIN 538 594 TSP TYPE-1 1.
FT DOMAIN 595 697 CYS-RICH.
FT DOMAIN 698 914 SPACER.
FT DOMAIN 915 990 TSP TYPE-1 2.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 997 AA; 109694 MW; 6587044ED02FC104 CRC64;

Query Match 30.7%; Score 237; DB 1; Length 997;
Best Local Similarity 18.2%; Pred. No. 0.026;
Matches 61; Conservative 24; Mismatches 62; Indels 188; Gaps 6;

OY 1 AAGGILHLEL-----LVAVG----- 15
DB 53 AGGSFLSYELWPRALKRQVSRDAPAFYELQYRGRELNFNLTAQHLLAPGFVSEFR 112
OY 16 -----PD-----VFOAHQED----- 25
DB 113 RGCLGRAHRAHTPACHLLGEVDPELEGGLAAISACDGLKGVQLSNEDYFIEPLDSAP 172
OY 26 ----- 25
DB 173 ARPGHAQPHVYKQAPERLAQRGDSAPSTCGVQVPELESRRERWQRQWRRLRR 232
OY 26 -----TERYVLTNLNGAELLRPDPSLGAQFRVHLV 55
DB 233 LHORSVSEKWKVETLVADAKMVEYHGQVQVESYVLTIMNVAGLPHDPSIGNPIHITV 292
OY 56 KMWILTEPEGAPNITANITSSLLSCVGSOTINPEDDTPGHADLVLTITFDELPLDGN 115
DB 293 RLVLLEDEEDLKHITHADNTLKSFCKWQKSNKMGDAHPHHTAILLTRKDL-CAAMN 351
OY 116 R--QVRGVTLGGACSPWSCLITEDTGFDLGVTI 148
DB 352 RPECETGLSHVAGMCQPHRSCSINEDTGLPLAFTV 386

RESULT 2
AT12_HUMAN STANDARD; PRT; 1593 AA.
AC PS0397;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-
DE TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RX MEDLINE=21264577; PubMed=11279086;
RA Cal S., Argueelles J.M., Fernandez P.L., Lopez-Otin C.;

```

```

RT "Identification, characterization, and intracellular processing of
RT ADAM-TS12, a novel human disintegrin with a complex structural
RT organization involving multiple thrombospondin-1 repeats.";
RL J. Biol. Chem. 276:17932-17940(2001).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
CC expressed in gastric carcinomas and in cancer cells of diverse
CC origin.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.
CC -1- TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1
CC DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AJ250725; CAC20419.1; -
DR Genem; HGNC:14605; ADAMTS12.
DR MIM; 606184; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprlysins.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTPeptdse.
DR Pfam; PF00090; tsp_1; 6.
DR Pfam; PF01421; Reprlysins; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 8.
DR PROSITE; PS00215; ADAM_MEPRO; 1. FALSE_NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS00092; TSP1; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 240 BY SIMILARITY.
FT CHAIN 241 1593 ADAMTS-12.
FT DOMAIN 465 544 DISINTEGRIN-LIKE.
FT DOMAIN 545 596 TSP TYPE-1 1.
FT DOMAIN 597 700 CYS-RICH.
FT DOMAIN 701 826 SPACER 1.
FT DOMAIN 827 881 TSP TYPE-1 2.
FT DOMAIN 886 943 TSP TYPE-1 3.
FT DOMAIN 947 995 TSP TYPE-1 4.
FT DOMAIN 996 1315 SPACER 2.
FT DOMAIN 1316 1364 TSP TYPE-1 5.
FT DOMAIN 1367 1423 TSP TYPE-1 6.
FT DOMAIN 1426 1471 TSP TYPE-1 7.
FT DOMAIN 1471 1471 TSP TYPE-1 8.
FT DOMAIN 302 305 POLY-GLU.
FT SITE 208 208 CYSTEINE SWITCH (POTENTIAL).
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 393 393 BY SIMILARITY.
FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 685 685 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 790 790 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 951 951 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1104 1104 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1275 1275 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1300 1300 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1320 1320 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 1503 1503 N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ SEQUENCE 1593 AA; 177545 MW; 07F9F48E63B83A3 CRC64;

Query Match 29.6%; Score 229; DB 1; Length 1593;
Best Local Similarity 18.3%; Pred. No. 0.26;
Matches 49; Conservative 39; Mismatches 61; Indels 123; Gaps 4;

Oy 3 GGILHLELLVAVGP----- 16
Db 124 GNLSHVKNMASSAPLCHLSGTVLQGGTRVGTAAALSACHGLTGFQPLPHGDFFTIEPVKKHP 183
Oy 17 -----DVFOAH----- 22
Db 184 LVGGYHPHIVYRQKVPETKCTGLKDSVNIQKQELWREKWRHNPRLSPSLRSRSIS 243
Oy 23 -----QEDTERVLTNLNIGAEILLRDPISLGAQFRVHLVXKMLTE 62
Db 244 KERWETLVVADTKMIEYHGSNVSILTIMNMVTLGPHNPISGNIAHIVVRLILLSE 303
Oy 63 PEGAPNITANLTSSLLSVCSOTINPEDTDPGHADLVLYITRFOLELPDGNR--QVRG 120
Db 304 EEGLKIVHAEKTLSSFCWKSINPKSDLPVPHVDVAVLLTRKDI-CAGFNRPCTELG 362
Oy 121 VTOLGGACSPWSCSLITETDGTGFDLGVTI 148
Db 363 LSHLSCWCPHRSNCINEDSGLPLFTI 390

RESULT 3.
EPPL HUMAN
ID EPPL HUMAN STANDARD; PRT; 5065 AA.
AC P58107;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epiplakin (450 kDa epidermal antigen).
GN EPIPL OR EPIPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=21201183; PubMed=11278896;
RA Fujiwara S., Takeo N., Otani Y., Parry D.A.D., Kunimatsu M., Lu R.,
RA Sasaki M., Matsuo N., Khaleduzzaman M., Yoshioka H.;
RT "Epiplakin, a novel member of the plakin family originally identified
RT as a 450-kDa human epidermal autoantigen: structure and tissue
RT localization.";
RL J. Biol. Chem. 276:13340-13347(2001).
CC -I- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST LEVELS IN LIVER,
CC SMALL INTESTINE, COLON, SALIVARY GLANDS, STOMACH AND APPENDIX.
CC -I- SIMILARITY: CONTAINS 65 PLECTIN REPEATS.
CC -I- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC DR Genew; HGNC:15577; EPPK1.
DR InterPro; IPR001101; Plectin_repeat.
DR Pfam; PF00681; Plectin; 36.
DR SMART; SM00250; PLEC; 65.
KW Coiled coil; Repeat; Structural protein; Cytoskeleton.
FT REPEAT 47 84 PLECTIN 1.
FT REPEAT 85 122 PLECTIN 2.
FT REPEAT 123 160 PLECTIN 3.
FT REPEAT 166 200 PLECTIN 4.
FT REPEAT 253 290 PLECTIN 5.
FT REPEAT 291 328 PLECTIN 6.
FT REPEAT 330 366 PLECTIN 7.
FT REPEAT 367 404 PLECTIN 8.
FT REPEAT 521 558 PLECTIN 9.
FT REPEAT 578 615 PLECTIN 10.
FT REPEAT 616 653 PLECTIN 11.
FT REPEAT 654 691 PLECTIN 12.
FT REPEAT 692 729 PLECTIN 13.
FT REPEAT 733 767 PLECTIN 14.
FT REPEAT 840 878 PLECTIN 15.
FT REPEAT 898 935 PLECTIN 16.
FT REPEAT 936 973 PLECTIN 17.
FT REPEAT 974 1011 PLECTIN 18.
FT REPEAT 1012 1049 PLECTIN 19.
FT REPEAT 1214 1251 PLECTIN 20.
FT REPEAT 1252 1289 PLECTIN 21.
FT REPEAT 1290 1327 PLECTIN 22.
FT REPEAT 1328 1365 PLECTIN 23.
FT REPEAT 1366 1403 PLECTIN 24.
FT REPEAT 1539 1576 PLECTIN 25.
FT REPEAT 1577 1614 PLECTIN 26.
FT REPEAT 1615 1652 PLECTIN 27.
FT REPEAT 1653 1690 PLECTIN 28.
FT REPEAT 1694 1728 PLECTIN 29.
FT REPEAT 1865 1902 PLECTIN 30.
FT REPEAT 1903 1940 PLECTIN 31.
FT REPEAT 1941 1978 PLECTIN 32.
FT REPEAT 1979 2016 PLECTIN 33.
FT REPEAT 2017 2054 PLECTIN 34.
FT REPEAT 2192 2234 PLECTIN 35.
FT REPEAT 2235 2272 PLECTIN 36.
FT REPEAT 2273 2310 PLECTIN 37.
FT REPEAT 2311 2348 PLECTIN 38.
FT REPEAT 2352 2386 PLECTIN 39.
FT REPEAT 2726 2768 PLECTIN 40.
FT REPEAT 2769 2806 PLECTIN 41.
FT REPEAT 2807 2844 PLECTIN 42.
FT REPEAT 2845 2882 PLECTIN 43.
FT REPEAT 2886 2920 PLECTIN 44.
FT REPEAT 3260 3302 PLECTIN 45.
FT REPEAT 3303 3340 PLECTIN 46.
FT REPEAT 3341 3378 PLECTIN 47.
FT REPEAT 3379 3416 PLECTIN 48.
FT REPEAT 3420 3454 PLECTIN 49.
FT REPEAT 3794 3836 PLECTIN 50.
FT REPEAT 3837 3874 PLECTIN 51.
FT REPEAT 3875 3912 PLECTIN 52.
FT REPEAT 3913 3950 PLECTIN 53.
FT REPEAT 3954 3988 PLECTIN 54.
FT REPEAT 4328 4370 PLECTIN 55.
FT REPEAT 4371 4408 PLECTIN 56.
FT REPEAT 4409 4446 PLECTIN 57.
FT REPEAT 4447 4484 PLECTIN 58.
FT REPEAT 4488 4522 PLECTIN 59.
FT REPEAT 4862 4904 PLECTIN 60.
FT REPEAT 4905 4942 PLECTIN 61.
FT REPEAT 4943 4980 PLECTIN 62.
FT REPEAT 4981 5018 PLECTIN 63.
FT REPEAT 5022 5056 PLECTIN 64.
FT REPEAT 1935 PLECTIN 65.
FT DOMAIN COILED COIL (POTENTIAL).

```

FT DOMAIN 2593 2624 COILED COIL (POTENTIAL) .
FT DOMAIN 3127 3158 COILED COIL (POTENTIAL) .
FT DOMAIN 3661 3692 COILED COIL (POTENTIAL) .
FT DOMAIN 4195 4226 COILED COIL (POTENTIAL) .
FT DOMAIN 4729 4760 COILED COIL (POTENTIAL) .
SQ SEQUENCE 5065 AA; 553078 MW; 378A472DEDA19489 CRC64;

Query Match 29.2%; Score 226; DB 1; Length 5065;
Best Local Similarity 1.9%; Pred. No. 11;
Matches 87; Conservative 25; Mismatches 36; Indels 4322; Gaps 26;

QY 1 AAGGI----- 5
| | |
Db 131 ATGGLVDPAGVVAPEPACHOGLLDRETHKLSLEPGTGLRFLNPTLRLTYHQLL 190
QY 6 ----- 5
| | |
Db 191 ERCVRAPGSLALLPLKITFRSMGAVSAAELLEVGILDEQAVOGLREGRLAAVDVSARA 250
QY 6 ----- 5
| | |
Db 251 EVRRYLEGTSVAGVVLPEGHKKSFFQOATEHLLPMGTALPLLEAQAATHLVDPITGQ 310
QY 6 ----- 13
| | |
Db 311 RLWDEAVRAGLVSPHELHEQLLVAEQAVTGHHPFSGSIPLFQAMKGLVDRPLALRL 370
QY 14 ----- 13
| | |
Db 371 DAQLATGVLCPARLRPLLEAALRCGLDEDTORQLSQAGSFGDTHGGURYSQALLALC 430
QY 14 ----- 13
| | |
Db 431 VTDPETGLAFPLSGPRGEPQPPFIKYSTRQALSTATATVSVKFGRPRVSLWELLF 490
QY 14 ----- 13
| | |
Db 491 SEAISSQRAMLAOYQEGTSLSEKLAELSATLEQAATAARVTFSGLRDVTVPGLLKA 550
QY 14 ----- 13
| | |
Db 551 EIIDQDLYERLEHQATAKDVGSLASAOYLQGTGCTAGLLPGSQERLSIYEARCKGLL 610
QY 14 ----- 18
| | |
Db 611 RPTALILLEAQAATGFIIDPKANKHSVEEALRAAVIGPDVFAKLLSAERAVTGYTDPY 670
QY 19 ----- 23
| | |
Db 671 TGOQISLFQAMKGLIVREHGIRLEAQAATGGVIDPVHSHRVPVDVAYRRGYFQDMLNL 730
QY 24 ----- 47
| | |
Db 731 ILLOPSDDTKGFFDNTNTHENTYLLQLLERCVRDPETGLYLLPLSSSTOSPLVDSATQAFQ 790
QY 48 ----- 47
| | |
Db 791 NLLLSVKYRFGQKVSAMELINSEYFSEGRRRQLRRYRQREVTLGQVAKLLEAETQRQ 850
QY 48 ----- 47
| | |
Db 851 ADIMLPALRSRVTHQLEAGIIDQQLDQVLAGTISPEALLMDGVRRYLCGLGAVGV 910
QY 48 ----- 47
| | |
Db 911 RLLPSGQRLSYQAMRQKLGPRVALLEAQAATGTIMDPHSPELSVDEAVRRGVGP 970
QY 48 ----- 51
| | |
Db 971 ELYGLKRAEAGIAGFRDPFSGKQSVFQAMKGLIPWEQAARLEAQAATGIIIDPTSH 1030
QY 52 ----- 57
| | |

Db 1031 HHLMPVAIQRGYVQDMETALSSSSETPFPDGOGRTSYAQLLBECPRDETSGLLHLLPL 1090
QY 58 V----- 58
| | |
Db 1091 PESAPALPTEEQVRSLOAVPGAKDGTSLMDLLSCHFTSEORRGLLEDVQEGRTTVPQL 1150
QY 59 ----- 58
| | |
Db 1151 LASVQRWVOETKLLAQAQAVMVPGPRGEVPVAVMLLDAGIITQETLEALAOGTSPAQVAEQ 1210
QY 59 ----- 66
| | |
Db 1211 PAVKACLWGTGCVAGVLLQPSGAKASIAQAVRDGLPTGLGQRLLEAQAASGFLVDPLNN 1270
QY 67 ----- 66
| | |
Db 1271 ORLSVEDAVKVLGVLGRELSEQLGQAEAAAAGYDPYVSRAASLSLWQAMEKGLVPQNEGLPL 1330
QY 67 ----- 66
| | |
Db 1331 LQVOLATGGVDPVHGHLPMQAAACRLGLLDTQTSOVLTAVDKDNKFFPDFPSARDQVYQ 1390
QY 67 ----- 66
| | |
Db 1391 QLRERCVCSETGLLLPLPSTVLEVDHTAVALKRMKVPVSTGRFKGCSVSLMDLLS 1450
QY 67 ----- 66
| | |
Db 1451 EYVGADKRRELVALCRSGRAAALRQVVSATVALVEAAERQPLQATFRGLRKQVSARDLPR 1510
QY 67 ----- 66
| | |
Db 1511 AQLISRKTLDELDSQTTTTVYKEVEMDSVKRSLEGGNFAGVLIQGTQERMSIPEALRRHI 1570
QY 67 ----- 66
| | |
Db 1571 LRPGTALVLEAQAATGFIIDPAENRKLTVBEAFKAGMFGKETYVKLLSAERAVTGYTDP 1630
QY 67 ----- 66
| | |
Db 1631 YTGQISLFQAMKGLIVREHGIRLEAQAATGGIIDPVHSHRVPVDVAYRCGYFDEMN 1690
QY 67 ----- 73
| | |
Db 1691 RILADPSDDTKGFFDNTNTHENTYLLQLLERCVRDPETGLYLLQIKKGENYVYINEATRH 1750
QY 74 ----- 73
| | |
Db 1751 VLQSR TAKMRVGRFADQVVSFMDLLSSPYFTEDRKRELIQYGAOSGGLEKLEIITTTI 1810
QY 74 ----- 73
| | |
Db 1811 EETETQNOGIKVAATRGVEVTAADLFNSRVIDOKTILTLRVGRTGGQALSTLECVKPYLEG 1870
QY 74 ----- 81
| | |
Db 1871 SDCIAGVTVPSTREVMSLHEASRKELIPAAAFATWLLAQAATGFLDPCCTROKLSVDEAV 1930
QY 82 ----- 81
| | |
Db 1931 DVGLVNEELRLKKAERATGYRDPATGDTIPLFQAMOKOLIEKAEALRLLEVOVATGG 1990
QY 82 ----- 81
| | |
Db 1991 VIDPOHHRLPLETAYRRGCLHKDIYALISDQKMRKRFVDPNTOEKVSYRELOQRCRPO 2050
QY 82 ---GW----- 83
| | |
Db 2051 EDTGWLPFVNKAARDSEHIDDETRALEAQEVEITVGRFRGQKPTLWALLNSEYVTEEK 2110
QY 84 ----- 87
| | |
Db 2111 KLQVLRMYRTHTRRALQTVAAQLILELIEKQETSNKHLWFOGIRROITASELLSSAIITEE 2170
| | |


```
QY 88 ----- 87
Db 2171 MLQDLTGSTTQELMEDDRVKRYLGSTSCIAGVLVPAKDQPGRQEKMSIYQAMWKGVL 2230
QY 88 ----- 87
Db 2231 PGTALVLEAQAATGFVIDPVRNRLSVVEEPVAGVVGSEIOEKLLSAERAVTGYTDPYT 2290
QY 88 ----- 87
Db 2291 GQOISLFOAMQKDLIVREHGIRLLEAQIATGGVIDPVHSHRPVDVAYRRGYFDEEMNRV 2350
QY 88 --NPEDDT----- 93
Db 2351 LADPSDTKGFPDNTNENLTVYVQLLRRCPDPDPTGLYMLQLAGRGSVAHQSEELRCAL 2410
QY 94 ----- 93
Db 2411 RDARVTPGSGALQGQSVSWELLFYREVSEDRRQDILLSRYRAGTLTVEELGATLTSLLAQ 2470
QY 94 ----- 93
Db 2471 AQAQARAEAGSPRDPREALRAATMEVKVGRGRVAVPVDVLASGYVSRAREELLA 2530
QY 94 -----DPG----- 96
Db 2531 EFGSGTLDLPALTRRLTAIEEAEAPGARPOLQDARRGPREPGPAGRGDGSRSQREG 2590
QY 97 ----- 96
Db 2591 QGEGETOEAATAARRQEQTLRDATMEVQGFQGRPVSVMDVLFSSYLSEARRDELL 2650
QY 97 --HA-----DLVLYITR----- 106
Db 2651 AQHAAGALGPLDVLAVLTVRIETEERLSKVSFRGLRQVSASELHTSGILGPETLRDLA 2710
QY 107 ----- 106
Db 2711 QGKTLOEVTEMDSVKRYLEGSTCIAGVLVPAKDQPGRQEKMSIYQAMWKGVL 2770
QY 107 ----- 106
Db 2771 LLEAQAATGFVIDPVRNRLSVVEAAGVVGSEIOEKLLSAERAVTGYTDPYTGQOISL 2830
QY 107 ----- 106
Db 2831 FOAMQKDLIVREHGIRLLEAQIATGGVIDPVHSHRPVDVAYRRGYFDEEMNRVLADPSD 2890
QY 107 ----FD----- 108
Db 2891 DTKGFPDNTNENLTVYVQLLRRCPDPDPTGLYMLQLAGRGSVAHQSEELRCALRDART 2950
QY 109 ----- 108
Db 2951 PGSGALQGQSVSWELLFYREVSEDRRQDILLSRYRAGTLTVEELGATLTSLLAQAQAAR 3010
QY 109 ----- 108
Db 3011 AEAEAGSPRDPREALRAATMEVKVGRGRVAVPVDVLASGYVSGAAREELLAEFSGT 3070
QY 109 LELP-----DGN----- 115
Db 3071 LDLPALTRRLTAIEEAEAPGARPOLQDARRGPREPGPAGRGDGSRSQREGQEGT 3130
QY 116 ----- 115
Db 3131 QEAAAAAAAAARQEQTLRDATMEVQGFQGRPVSVMDVLFSSYLSEARRDELLAQAHAAG 3190
QY 116 -----ROV----- 118
Db 3191 ALGPLDVLAVLTVRIETEERLSKVSFRGLRQVSASELHTSGILGPETLRDLAOGTKTL 3250
```

```
QY 119 ----- 118
Db 3251 QEVTEMDSVKRYLEGSTCIAGVLVPAKDQPGRQEKMSIYQAMWKGVL 3310
QY 119 -----RGVT----- 122
Db 3311 ATGFVIDPVRNRLSVVEAAGVVGSEIOEKLLSAERAVTGYTDPYTGQOISLFOAMQK 3370
QY 123 ----- 122
Db 3371 DLIVREHGIRLLEAQIATGGVIDPVHSHRPVDVAYRRGYFDEEMNRVLADPSDDTKGFF 3430
QY 123 ----- 122
Db 3431 DPNTNENLTVYVQLLRRCPDPDPTGLYMLQLAGRGSVAHQSEELRCALRDARTVPGSAL 3490
QY 123 ----- 122
Db 3491 QGQSVSWELLFYREVSEDRRQDILLSRYRAGTLTVEELGATLTSLLAQAQARAEAG 3550
QY 123 -----QLGGACSPW----- 132
Db 3551 SPRDPREALRAATMEVKVGRGRVAVPVDVLASGYVSGAAREELLAEFSGTLDLPAL 3610
QY 133 ----- 132
Db 3611 TRRLTAIEEAEAPGARPOLQDARRGPREPGPAGRGDGSRSQREGQEGETOEAATA 3670
QY 133 ----- 132
Db 3671 AAAARRQEQTLRDATMEVQGFQGRPVSVMDVLFSSYLSEARRDELLAQAHAAGALGPLD 3730
QY 133 ----- 132
Db 3731 LVAVLTVRIETEERLSKVSFRGLRQVSASELHTSGILGPETLRDLAOGTKTLQEVTEM 3790
QY 133 -----SCL----- 135
Db 3791 DSVKRYLEGSTCIAGVLVPAKDQPGRQEKMSIYQAMWKGVL 3850
QY 136 ----- 135
Db 3851 DPVRNRLSVVEAAGVVGSEIOEKLLSAERAVTGYTDPYTGQOISLFOAMQKDLIVRE 3910
QY 136 ----- 135
Db 3911 HGIRLLEAQIATGGVIDPVHSHRPVDVAYRRGYFDEEMNRVLADPSDDTKGFPDNTHE 3970
QY 136 ----- 135
Db 3971 NLTVYVQLLRRCPDPDPTGLYMLQLAGRGSVAHQSEELRCALRDARTVPGSGALQGQSV 4030
QY 136 -----ITED----- 139
Db 4031 VWELLFYREVSEDRRQDILLSRYRAGTLTVEELGATLTSLLAQAQARAEAGSPRDP 4090
QY 140 ----- 139
Db 4091 REALRAATMEVKVGRGRVAVPVDVLASGYVSRAREELLAEFSGTLDLPALTRRLTA 4150
QY 140 ----- 139
Db 4151 IIEEAEAPGARPOLQDARRGPREPGPAGRGDGSRSQREGQEGETOEAATAAARR 4210
QY 140 ----- 139
Db 4211 QEQTLRDATMEVQGFQGRPVSVMDVLFSSYLSEARRDELLAQAHAAGALGPLDVAVUT 4270
QY 140 ----- 139
Db 4271 RVIEETEERLSKVSFRGLRQVSASELHTSGILGPETLRDLAOGTKTLQEVTEMDSVKRY 4330
QY 140 -----TGF----- 142
```

```

Db 4331 LEGTSCIAGVLVPAKQDPGRQEKXSIYQAMWKGVLRLPGTALVLEAQAATGVIDPVRNL 4390
Qy 143 -----142
Db 4391 RLSVEEAVAGVVGGEIOEKLLSNERAVTGYDPTGQQISLFQAMQKDLIVREHGIRLL 4450
Qy 143 -----142
Db 4451 EAQIATGGVIDPVHSHRPVVDVAYRGYFDEEMNRVLADPSDDTKGFPDNTHTNYVQ 4510
Qy 143 -----142
Db 4511 LLRCVDPDPTGLYMLQAGRSNAVHOLSELRCALRDARVTPGSGALQSGSVSWELLF 4570
Qy 143 -----DLGVTI 148
Db 4571 YREVSERODLLSRYRAGTLTVEELGATL 4600

RESULT 4
BACA_BACLI STANDARD; PRT; 5255 AA.
AC 068006;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Bacitracin synthetase 1 (BAI) [includes: ATP-dependent isoleucine
DE adenylation (ileA) (isoleucine activase); ATP-dependent cysteine
DE adenylation (cysA) (cysteine activase); ATP-dependent leucine adenylation
DE (leuA) (leucine activase); ATP-dependent glutamate adenylation (gluA)
DE (glutamate activase); ATP-dependent isoleucine adenylation (ileA)
DE (isoleucine activase); Glutamate racemase (EC 5.1.1.3)].
GN BACA.
OS Bacillus licheniformis.
OC Bacteria, Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10716;
RX MEDLINE=98089193; PubMed=9427658;
RA Konz D., Klens A., Schoerendorfer K., Marahiel M.A.;
RT "The bacitracin biosynthesis operon of Bacillus licheniformis ATCC
RT 10716: molecular characterization of three multi-modular peptide
RT synthetases."
RL Chem. Biol. 4:927-937(1997).
CC -1- FUNCTION: ACTIVATES FIVE AMINO ACIDS, INCORPORATES TWO D-AMINO
CC ACIDS, RELEASES AND CYCLIZES THE MATURE BACITRACIN.
CC -1- CATALYTIC ACTIVITY: L-glutamate = D-glutamate.
CC -1- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC (POTENTIAL).
CC -1- PATHWAY: Cyclic peptide antibiotic bacitracin biosynthesis.
CC -1- SUBUNIT: LARGE MULTIMERIC COMPLEX OF BAI, BAI2 AND BAI3.
CC -1- DOMAIN: CONSISTS OF FIVE MODULES AND ONE EPIMERIZATION DOMAIN IN
CC THE FOURTH MODULE. EACH MODULE INCORPORATES ONE AMINO ACID INTO
CC THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS
CC RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOLATION, CONDENSATION
CC (NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND
CC N METHYLATION (OPTIONAL).
CC -1- MISCELLANEOUS: BACITRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC
CC DODECAPEPTIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST
CC ABUNDANT IS BACITRACIN A, A BRANCHED CYCLIC DODECAPEPTIDE. IT
CC CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-
CC GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAZOLINE CONDENSATION
CC PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-
CC HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-
CC TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT
CC CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,
CC PHE-9, AND ASP-11).
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF007865; AAC06346.1; -.
DR HSSP; P14687; 1AMU.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Ppantne attach.
DR Pfam; PF00501; AMP-binding; 5.
DR Pfam; PF00550; pp-binding; 5.
DR Pfam; PF00668; Condensation; 5.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR PROSITE; PS00455; AMP BINDING; 5.
DR PROSITE; PS50075; ACP_DOMAIN; 5.
DR Ligase; Isomerase; Antibiotic biosynthesis; Phosphopantetheine;
KW Multifunctional enzyme; Repeat.
FT REPEAT 39 612 DOMAIN 1 (ISOLEUCINE-ACTIVATING).
FT REPEAT 1109 1648 DOMAIN 2 (CYSTEINE-ACTIVATING).
FT REPEAT 2124 2689 DOMAIN 3 (LEUCINE-ACTIVATING).
FT REPEAT 3164 3732 DOMAIN 4 (GLUTAMINE-ACTIVATING).
FT REPEAT 4668 5249 DOMAIN 5 (ISOLEUCINE-ACTIVATING).
FT DOMAIN 621 1037 CYCLIZATION (POTENTIAL).
FT DOMAIN 544 611 ACYL CARRIER (ACP) 1.
FT DOMAIN 1585 1652 ACYL CARRIER (ACP) 2.
FT DOMAIN 2621 2888 ACYL CARRIER (ACP) 3.
FT DOMAIN 3664 3730 ACYL CARRIER (ACP) 4.
FT DOMAIN 5171 5238 ACYL CARRIER (ACP) 5.
FT BINDING 574 574 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 1615 1615 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 2651 2651 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 3694 3694 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 5201 5201 PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ SEQUENCE 5255 AA; 598254 MW; 906B8DD6450F95B CRC64;

Query Match 29.0%; Score 224; DB 1; Length 5255;
Best Local Similarity 1.6%; Pred. No. 15;
Matches 77; Conservative 34; Mismatches 36; Indels 4569; Gaps 24;

Qy 2 AGG-----4
Db 110 AGGAYLPIDPETPKDRIAPFMSDTRAAVLLTQCKAAGIDCEADIVQLDREASDGFSEK 169
Qy 5 -----4
Db 170 LGSVNDSDGTAVIITSGTGTGPKGVITPHYSVIRVVQNTNIDITEDNVILQLSNYSFD 229
Qy 5 -----4
Db 230 GSVDFIFGALLNGASLVMIKEALLNIRLGSAINEEKVSMFITTALFNMIADIVDCL 289
Qy 5 -----ILHL-----8
Db 290 SNLRKILFGGERASIPHRVKVNLHVGRDKLIHVYGPTESTVTVATYTYFNEIDDEAETIPI 349
Qy 9 -----8
Db 350 GSPLANTSVLIMDEAGKLVPIGVPGELCIAGDGLSGVLYNREELTAETKFIHPHFIPGERL 409
Qy 9 -----8
Db 410 YKTGDLAKWLPDGNIEFIGRIDHGVKIRGFRIELGIESRLEMHEDINETIVTVREDEES 469
Qy 9 -----8
Db 470 RPYICAVITANREISLDELKGLGELKPEYMIPIYFVKLDKLPKTKNGKVDKALPEPDR 529
Qy 9 -----ELLVAVGPDVFAQ-----21

```

Db 530 TAGAENEYAPRNETEEKLAAMVQGVLVHVEKAGIHDFHFAQMGHSLHAMELIAKIKKXN 589
QY 22 -----HOED----- 25
Db 590 VEIPLHQLFKLATIKELSAFIENHOEDGDTLVTRAADPENIHEIFPLTGTQLAYLVGR 649
QY 26 -----TER-- 28
Db 650 DETFEIGGVATNLTVFEADVDLNRFLTQKLDIDRHPILRTIVFENGOKILEATQRT 709
QY 29 ----- 28
Db 710 IETQDLRGFTBEENVRILEOREKMTSKIIDPSWPLFELKTFMLPGEKKYFFLNVDPLI 769
QY 29 -----YVLTNLNI----- 36
Db 770 CDDSSMKRLIREFKQLYENPGLQPSLEYSFRDYVLASINFKQTSRYQKXQDQWLDKLDH 829
QY 37 ----- 36
Db 830 FPSAPELPLKSDPAHVAKPSFKFSTFLDGHWTWNLKELKKARHHHLTPTSVLCAAYAYILA 889
QY 37 ----- 36
Db 890 YMSRONHFALNTVFNRIIPFHPDKVMIGDFTSLMLLDIHAENMSSFWRFALNVQDTLL 949
QY 37 -----GAELLR----- 42
Db 950 EALEHRRYDGVDIRNIAKNGKNKKAAMPITVTSVLSENPDSDSLVDNFONIHFFSTR 1009
QY 43 ----- 42
Db 1010 TSQVYIDNQVEINGLYITWDYVEQI FEHEVIESMFDQYIAVIQKAVSGEDVSTIQMNE 1069
QY 43 ----- 42
Db 1070 KSRQIMISAYNDTQSDPAKPLHELFTGOVKGHPDRMAKHHDEWMTYOELDEKSNQVAF 1129
QY 43 -----DP----- 44
Db 1130 LIGKGEKGDYIGVIGKRSGLTIVNLLAVLKTAAGAVIPLDPDYPERKAYIQSKSNCKFF 1189
QY 45 ----- 44
Db 1190 ISHDVYDKEHIERFSKAPVDRKVDLDDMAVIFTSGTGKPGVQITHVPQRNTILDINE 1249
QY 45 -----SLGAQFRV----- 52
Db 1250 KFNVTQDNIMGSSLCFDSLVSVDVFCALSSGASLVIIDQDRDVFSLKETAERITIW 1309
QY 53 -----HLVKMWIL----- 60
Db 1310 SVPAIMGMTADVDPDNELNHLRLIILSGDWIPLQLPATIKKTFKNAEVISLGGATEGSI 1369
QY 61 ----- 60
Db 1370 WSIYPIQKVEEDKSI PYGKPLANQKIYVLNQNKLCPVGVEGELYIGGAVASGYIHD 1429
QY 61 ----- 60
Db 1430 OEKTEHSFIOHOBELGYIYKTGDYVLKEDGYVEFLGRKDSQVKIRGYRVEMGIENTLVS 1489
QY 61 -----TEPEGAPNITA----- 71
Db 1490 HOEITKASVIDYTPDGIKNLYAFVVAENAI SLDVKEFLQKTLPDYMI PAKFVQIEEIP 1549
QY 72 ----- 71
Db 1550 LTVNGKVDKRTLHLAEQHTADEQGRMLPENETQAMLEIWKDIFGLDSINLDVSY 1609
QY 72 ----- 71

Db 1610 EIGDSLKASIIITEINKRMVEMPISFIFKNDTIILADHLYLKNREESDMEHPIQKAREK 1669
QY 72 ----- 71
Db 1670 EYPTSPAQRMYMLSMLENERGAYHIPALLVEGRINAMQLENALKTFLOREILRTGF 1729
QY 72 -----NLTS----- 75
Db 1730 EIONNELIQIYENVDFRLEYECLDASITDOHALMEITSRYCKESIKFPDLSRPLMRAK 1789
QY 76 ----- 75
Db 1790 LIKIDDIRHILVINPHHII SDGVSQILMNEILELNSVPLPEVNVQYKDYVEMHHTFNQ 1849
QY 76 ----- 75
Db 1850 SAAMKKQAYWLDVYRIDPSKLDPPDYKRHHIDTFEGSSVPLEMERELSDHIRKLAKHN 1909
QY 76 ----- 75
Db 1910 GTTLYTVMLSAYYYVLLNKYTNQTDIVGTAAGRHLHPLQDVGFGVNTLALRNEVDTSY 1969
QY 76 ----- 75
Db 1970 SFKEFLOQTERTIAAFDNSEYPPDDLLIRKLVGRESNRNPLDFTMFVLEDARMTKQKG 2029
QY 76 ----- 75
Db 2030 DVKLSPIIFELDNAKFDIMFNVLDFEQIKVLNIEYSTLSFKDQETIQKIAEDYFRILEEVS 2089
QY 76 ----- 75
Db 2090 ENLDVALHQIDMISROEKRTLLESFNHTKTAYPKGKAHQLFEEQAKRIPDHTAVVFEDQ 2149
QY 76 ----- 75
Db 2150 KLTYQLNEKANQVALLREKGVKPDTLVGIMMERSSDMAAILGLVKAGGAYLPIDPEY 2209
QY 76 ----- 75
Db 2210 PPERMYAFDSEVKVVIISDVPLAELTAESIELIHMDDERIAGQDRSDIDNVNOSGDLA 2269
QY 76 -----SLLSVCWHSQ----- 85
Db 2270 VVIYTGSGTKPGYMIHQSLINLCSWHQSCFEVQNDNSSIYASISFDPAPVWELFPYI 2329
QY 86 ----- 85
Db 2330 TAGATVHVNLNOETRLDVEKLNRYFHDHHTISFLPTPVEQFTALDNLHSLRLLTGGDKL 2389
QY 86 ----- 85
Db 2390 NVFKEKSYQIVNNYGPTEVTVVATSPIDKSHQNIPIGKPIDNVKVYILNKDQLCLPGA 2449
QY 86 ----- 85
Db 2450 SGELCIAGEGLARGYVNRPELTREKFIGNFPVPGRMRYRTGLAKMLPDGNIQFLGRVDQ 2509
QY 86 ----- 85
Db 2510 QVKIRGYRIEPEIEIENRLLKYEIEEAAVIAREDGDHDPYLCAYTVVKKEVEPEKIRAF 2569
QY 86 -----TIN-----PBD----- 91
Db 2570 KKSPLYMIPOYFVQLDGLPLTVNGKVDKSLPVPERSVTMDRRYEAPRDOMEEKLVSIW 2629
QY 92 ----- 91
Db 2630 EEPALGINKIGINSHFPEAGHSLKAAALVSTIHKELNVKLPLURQIFETPTIKGLRDISVR 2689
QY 92 ----- 91
Db 2690 RRCFCYIDRTEEKPYRSLSSAQKELYILSQTGSHVAYNMPFAMTLEGDFDIRRENTLUK 2749

```
QY 92 ----- 91
Db 2750 NMVKHESFRTFWMIDGVNQIEKIDFOVAYSDIGKSAEBKISFIRPFHLEKAPL 2809
QY 92 ----- 91
Db 2810 LRAEVVKLNEREHLMLFDMHHIISDGVSTDFIQELGALYEGKSLKPFPHIQYDYAEWEN 2869
QY 92 ----- 91
Db 2870 SHARSEELKQOEYWLTKYKGDIPVLDLPIDHKRPLTKSSEGDVTAAIESETPRKLQHM 2929
QY 92 ----- 91
Db 2930 AKENGVTMYMLLAGYTALLSKYTGOEDIIVGTTPAAGRHNHEDIOHLIGMFVNTLAIRNHP 2989
QY 92 ----- 91
Db 2990 EGKKTFRDYLOEVKENTLOAYENODYPFEELVEKVNIKRDMARNPLFDTMLVYHNTDVKP 3049
QY 92 ----- 91
Db 3050 FEAGLSRLVEIKRGISKEDITVTASEAADGLRLEVEYSTTLFNKERMERLSEHLISLL 3109
QY 92 ----- 91
Db 3110 EQAADHPDIAINOIDVLTGERHRVLYDNRTDGVFCCKEMTIPELFEKQAEKTPDHPAVA 3169
QY 92 ----- 91
Db 3170 FGDETI SYRELNERANSLAFTLROKGVGPDVIAGILTERSIEMIVGIMILKAGGAYLPI 3229
QY 92 ----- 96
Db 3230 DPAYQERISYIVKSDVSVCAGDVPGEAYTGDIIRIDQTQONDHVENLKHDIKPOH 3289
QY 97 ----- 96
Db 3290 LAYVIYTGSTGKPKGVMIHHSVNNLVHGLNERIYOHLDALNVALVAPYIFDASVKQI 3349
QY 97 ----- 110
Db 3350 PAALLFGHTLCIVPRETAWDAMSLIEYYSKNNINVSMTPAHLNMLAYVDKTELEFDVKE 3409
QY 111 ----- 110
Db 3410 LIVGDDALTPDVIIGLPHKFPNLSCNITNVYGPTECCVDAASHQIESGKVPQTPSIPIGR 3469
QY 111 ----- 110
Db 3470 PLLNTSIYIVDKELRPLPVGIAGELCIAGEGVARGVNNRPELTAEKFVDHPFEPKXMYK 3529
QY 111 ----- 114
Db 3530 TGDMLMLPDCQIEFLGRADHQVKIRGYRIELGEVEQQLTHEKIKEAAVITAGKDQNGS 3589
QY 115 ----- 114
Db 3590 YLCAYIASDKELPAADVQFLEREMPDYMIPSYFVKLORLPRTPSGKVDRSALPEADGNV 3649
QY 115 ----- 114
Db 3650 NVMEGTGYDPPRNEIERKLQVWREILGAEDIGISHHFFAAGGDSIKALQIVSRKMNML 3709
QY 115 ----- 124
Db 3710 KLEMKALFANPKIKDLSRFITEETHRKHKNKPVGTETELLPIQKRYFANNKEELDHFNQS 3769
QY 125 ----- 124
Db 3770 FMLFRKGDYDENIVPTAFNKILEQHDALRMIYEKDDGDIQVNRGYRENLFOLDVYDVRG 3829
```

```
QY 125 ----- 124
Db 3830 FDSOEKVPFELATGIOKSSIRKGLVHLGI1PRADEGDHLLIAIHHLVWGVDSWRILFED 3889
QY 125 ----- 124
Db 3890 FETLYQALKGEPLDIGYKTDYSOEFAQLKKYQAQRLLKEREYWKALEADVFPFPAE 3949
QY 125 ----- 124
Db 3950 KLERDFFHSATLSIRIGPDVTAKLLRNAPKAYNTEINDILLTALIAAVRDIITGENKLV 4009
QY 125 ----- 124
Db 4010 MMEGHGREDILGVDITRTIGWFTTVVPFIDLGBEKEISONIKWVKEALRKIPNKGIGY 4069
QY 125 ----- 124
Db 4070 GVLKYMTEELQKIQTOAPLSFNFGEMNNDMNRKVSQSPSPGESIGGKIVRHCAIEMN 4129
QY 125 ----- 131
Db 4130 AISLNGELTIYTFNODQYOTSTIEOLNOSFKENLEKIVDHCVDKESDMTSPDYGDVSL 4189
QY 132 ----- 131
Db 4190 GLSELELIKDYSAFOIEKIYPLANMOKGMLPHNAMDQTSQAYFQQIVIKLGRVHPDIL 4249
QY 132 ----- 131
Db 4250 EBSFHEIVKHEILRASFEYEITAEPRQIIARDKTPFTSIDLTGENRTRQHRPIETYLK 4309
QY 132 ----- 133
Db 4310 EDOEKGPDLSSEALMRVCLIKMSDESRYRLIWSHHILLDWCGLGIVLSELSYLGKIMKG 4369
QY 134 ----- 134
Db 4370 ESRLKEPKPYGDYIKWLEKQDOEAVAYWKDYLGYESRSELPAPNRGATSEECCKEK 4429
QY 135 ----- 134
Db 4430 VISFSKELTTKTRIAKOHHVTINTVLQIGWIMILAKYKNTDEVVFGTVVSGREAPVDGI 4489
QY 135 ----- 137
Db 4490 EEMVGLFIHTPTTRISFEGARSFKEVLKKTQAESIESNRYSYNNLSIEIQLSEMKRELIT 4549
QY 138 E-----DTGFDL----- 144
Db 4550 HVMAFQNYAFDEELFRSQSGETGFELEGVHGKERTNYNFNLTGVLEDEQLKLTFFENV 4609
QY 145 ----- 144
Db 4610 YDNTIETLEKHIITVAEQVAEDETQTLRDINLVSKEEORHLIHTFNDTKTGYPKDKPLH 4669
QY 145 ----- 144
Db 4670 ELFEEQAMKTPDHTALVFGAORMTYRELNEKANQATARLLREKIGRGSIAAIIADRSFEM 4729
QY 145 ----- 144
Db 4730 IIGIIGILKAGTYLPIDPETPRDRIDYMLKNSGAALLVTTDSLKLPFDIKTVDLCSDEL 4789
QY 145 ----- 148
Db 4790 HLLSEENLPRNRSDDTAYIVYTSGSTGTPKGVVI 4824
```

```
RESULT 5
HTS1 COCCA
ID HTS1 COCCA
AC Q01866;
PRT; 5217 AA.
```

DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE HC-toxin synthetase (EC 6.3.2.-) (HTS).
GN HTS1.
OS Cochliobolus carbonum (Bipolaris zeicola).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; Cochliobolus.
OX NCBI_TaxID=5017;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 90305 / SB111;
RX MEDLINE=93100328; PubMed=1281482;
RA Scott-Craig J.S.; Panaccione D.G.; Pocard J.-A.; Walton J.D.;
RT "The cyclic peptide synthetase catalyzing HC-toxin production in the
RT filamentous fungus Cochliobolus carbonum is encoded by a
RT 15.7-kilobase open reading frame.";
RT J. Biol. Chem. 267:26044-26049 (1992).
RN [2]
RN FUNCTION.
RP STRAIN=ATCC 90305 / SB111;
RX MEDLINE=20138231; PubMed=10671527;
RA Cheng Y.-Q.; Walton J.D.;
RT "A eukaryotic alanine racemase gene involved in cyclic peptide
RT biosynthesis.";
RL J. Biol. Chem. 275:4906-4911 (2000).
CC -!- FUNCTION: Non-ribosomal peptide synthetase, able to activate
CC proline and ABO (2-amino-9,10-epoxy-8-oxodecanoic acid), and
CC epimerize L-Pro. Catalyzes the production of HC-toxin: a cyclic
CC tetrapeptide. Activates and thioesterifies L-Pro, and epimerizes
CC it to D-Pro; also uses D-Ala as a substrate but this is epimerized
CC from L-Ala by TOXG.
CC -!- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF HC-TOXIN.
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -!- SIMILARITY: CONTAINS 4 ACYL CARRIER DOMAINS.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M98024; AAA33023.1; -;
DR HSP; P14687; 1AMU.
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR001242; Condensatn.
DR InterPro: IPR003880; Ppantne_attach.
DR Pfam: PF00501; AMP-binding; 4.
DR Pfam: PF00550; pp-binding; 4.
DR Pfam: PF00668; Condensation; 5.
DR PRINTS: PR00154; AMPBINDING.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 4.
DR PROSITE: PS00455; AMP BINDING; 3.
DR PROSITE: PS00075; ACP DOMAIN; 4.
KW Ligase; Multifunctional enzyme; Phosphopantetheine; Repeat.
FT REPEAT 249 842 DOMAIN 1.
FT REPEAT 1854 2452 DOMAIN 2.
FT REPEAT 3006 3606 DOMAIN 3.
FT REPEAT 4158 4738 DOMAIN 4.
FT DOMAIN 762 840 ACYL CARRIER (ACP) 1.
FT DOMAIN 2384 2450 ACYL CARRIER (ACP) 2.
FT DOMAIN 3536 3604 ACYL CARRIER (ACP) 3.
FT DOMAIN 4667 4736 ACYL CARRIER (ACP) 4.
FT BINDING 803 803 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 2414 2414 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 3568 3568 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 4700 4700 PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ SEQUENCE 5217 AA; 574637 MW; 0331D9C5400163A5 CRC64;

Query Match 28.7%; Score 222; DB 1; Length 5217;
Best Local Similarity 1.8%; Pred. No. 19;
Matches 87; Conservative 20; Mismatches 36; Indels 4702; Gaps 27;
QY 2 AGGI----- 5
DB 289 AGGVCVPIDPRYPVERIRDIIRTNATIALVGAGKTAALFKSADTAVQITIDITKDIPHGL 348
QY 6 ----- 5
DB 349 SDTVQSNKIDDPAPGLFTSGSTGVPKCIIVVTHSOICTAVQAYKDRFGVTSETRVLQFS 408
QY 6 ----- 5
DB 409 SYTFDISIADTFTALFYGGTLCIPSEEDRMSNLQDYMVSVRPNWAVLTPTVSRFLDPGVV 468
QY 6 ----- 5
DB 469 KDFISTLIFTGEASREADTPWIEAGVNLVNYGPAENTLITTATIRKGSNSNIGYVN 528
QY 6 ----- 5
DB 529 TRTWTVDSGACLVPGSIGELLIESGHLADKYLNRDPRTAAFLSDLPMIPLYEGDSVR 588
QY 6 ----- 5
DB 589 RGRPRYRTGDLVRYCDDGSLICVGRSDTQIKLAGORVELGDVEAHLQSDPTTSQAAVFP 648
QY 6 ----- 5
DB 649 RSGPLEARLIALLVTGNKDGTPHNOQLPKAPAFAPCPDLVKYATSSLOORLPYSYMPVS 708
QY 6 ----- 5
DB 709 WLGDPLPMSVSGKLDRAVLQDQLESFSDYAEILGTTGLEVDPGGAASSVASDSDLRD 768
QY 6 ----- 5
DB 769 MNDDSLLTACSRVNLNLPAGKISYSQSFTIHAGDSITAMQVSSWMKRFPTGKIGVKDLLV 828
QY 6 ----- 5
DB 829 SPSISTAASCIKSAQDGSRNFAVRPGORIPVSPICKLFFQTAEASKSNHMYHQSFPR 888
QY 6 ----- 5
DB 889 DQPIKPTQIEDAISLMORHPMLQARFETEGDWYQYIPIDVERRASVEVIGSLSTDDR 948
QY 6 ----- 5
DB 949 EAAMLRARQSIDLTEGLIRCOLFNNNVDEASRLFFVVIHHAVIDLVSVRIIMEELEAHL 1008
QY 6 ----- 5
DB 1009 ATDSTPDRGEAYQESVPFLAWCQVQAEAVKDIPTDRTVPLIPKIPADFGYGLKHENV 1068
QY 6 ----- 7
DB 1069 YGNTVERKIPLGHSITEDLLYKCHDSLHTKTTIDVLLAAVLVSFRKSFLLDRPVPVFN 1128
QY 8 ----- 7
DB 1129 GREPGGEDAVLSRTVGWFTTISPVVPEVSPGDILDVVRVKDYRWATPNNGDFYFSTK 1188
QY 8 ----- 8
DB 1189 YLTQSGIKLFDHLPAEILFNVEGRYQAMESQTVLKPESWHAGEASKQDQDGLRRFCLF 1248
QY 9 ELLVAVGPD----- 17
DB 1249 EISTAVLPDQGLHLTCSMNKNRHOGRIRLWLDITLLPAAIGSIVSSLALASPOLTLS 1308

QY	18	-----	17
Db	1309	LLRLDYSSLDILKKSILSIPAVQTLDDLEGVPGSPQDQALFLSQSKSDGAYEVDFTW	1368
QY	18	-----VFOAQOED--	25
Db	1369	RVATSLQNSQPAVDIGCLVEANKDVALHAALRTVILESSLPATGILHQVVLRSRHDPIV	1428
QY	26	-----TE-----	27
Db	1429	ILDVRDVTAAITILDSYPPPTTEEGIALIKRPPHRLLICTTIEGSLVLIKFOVNHVLFQDMS	1488
QY	28	-----RYV-----	30
Db	1489	TDKIIQDLSKAYTCRHSNKLPHQESKLHDGTGNRPETKPLAEFIRYIRDPORKQDSIN	1548
QY	31	-----LTNLNIG--	37
Db	1549	YWKNALRGATTCSPFPLFDQITSEKAMPQSWASVPIPLCVDSKELSKTLANLGITSTM	1608
QY	38	-----	37
Db	1609	FQTVNAIVLRIYSONGQSVFGLTSGRDPAPVDGIDSAVGNFIAMLCVCFDFDDGQVHTA	1668
QY	38	-----	37
Db	1669	DMARKIHNASANSISHOASLAEIQDALGLSTSTPLENTAFTYLPKRPNTYKAGEPEHL	1728
QY	38	-----AELLR 42	
Db	1729	CFEELSMSDPTFDTLTFVEPTQESNEVSAHLDFKLKLSIQAYATSIASITVAHILSELVH	1788
QY	43	DP-----	44
Db	1789	DPYRALNTLPVSEHDTAIRSNDHLFPPEATECHETFSRKVVEHPQREAIKSWDGSLT	1848
QY	45	-----SLGAQFRVHLVKM-----	57
Db	1849	YAEISDLSQLSLHLSVLGKIVGTKIPICFEKSMWITVILAVQAGGVFVLEPQHPES	1908
QY	58	-----VILTEPE-----	64
Db	1909	RLSGIHKOVQAEILLCSPATSRMGALQNIQTGTEFKIVELEPEFIRSLPLPPKNHPQ	1968
QY	65	-----GAP-----	67
Db	1969	MVGLNDLYVVFTSGTGVKGAVALHOAYATGIVEHAVACGMTSLGAPPRSLQFASYSF	2028
QY	68	-----NITANL-----	73
Db	2029	DASIGDIFTTLAVGGCLCPREEDRNPAGITTFINRYGVTWAGITPSSLALHLDPAVPTL	2088
QY	74	-----	73
Db	2089	KALCVAGEPLSMVTVVWSKRLNLINMYGPTATVACIANQVCTTTTIVSDIGRCYRATT	2148
QY	74	-----	73
Db	2149	WVQPDNHNLSVPIGAVGELIIEGSLCRGYLNDPERTAEVFIRSPSLHDLRPNSTLYK	2208
QY	74	-----	73
Db	2209	TGDLVRSADGKIIFIQRKDTQVXKNGQRFELGEVEHALQLOLDFSDGPIIVDLLKRTQS	2268
QY	74	-----	73
Db	2269	GEPDLLIAFLVGRANTGTGNSDEIFIATSTSLSEFSTVIKKLODAQRAVEVLPFMVP	2328
QY	74	-----TSSLLS-----	79
Db	2329	QAYPIEGGIPLTAAGKIDRRMLRKLCEPFRNDLISFTSKALSTSVKDAETTTVDREL	2388
QY	80	-----	79

Db	2389	ARIWEKVLGVKVGRESDFSSGNGNSMAAIALRAEAQRSGFTLFVADIITPNRLADMAKL	2448
QY	80	-----	79
Db	2449	FSHQSVSPSSSTLRTKVPISLQKRSSGLOTAAPVSNGPSVRRCKENIIDCPVAFEYE	2508
QY	80	-----VCG-----	82
Db	2509	EGPSDTOLKEASRICGISRSIEDVPPCTPMQOEALVALSLIPGAQASALHAFAELRPG	2568
QY	83	-----	82
Db	2569	DRNFRSAMESVYKAQPIILRSRIISGSSGVVVTSDISIPOLDVSGLDTFLEOQLOVG	2628
QY	83	-----WSQ-----	85
Db	2629	FAPGAPLFLAFVYSKADCDYFVISAHHAIYDGNLSNLIWSQVLAALYTNGLPPLPGP	2688
QY	86	-----	85
Db	2689	XHFARNLNVQSKLDSDFWRKLLVKPDQESFRFPDVPVGHKPACTRTTNFHFPSMQSK	2748
QY	86	-----TIN-----	88
Db	2749	IGTANTCINAAWAILTAQYSSNKTVNFVGTWGRDPPMIDIEMTGPTIVTVPROVNI	2808
QY	89	-----	88
Db	2809	PESSVAEFLQDLQSLAVVLPHQHLGHRTOALGPRIARQACDFSTLLVNVHGSISWSEL	2868
QY	89	-----	88
Db	2869	EAADIVPVPLRSSDLYAYPMVEVENASSDTDIRVHSDPDCIEVOLLERLMEQFGHNLQ	2928
QY	89	-----	88
Db	2929	TLCRASDFPGKRIAEMLDDTATTTLRLTFLSMNSRVKSDPDVAAIAVHKLEETAOSQA	2988
QY	89	-----	88
Db	2989	ESAIVAHGOLSYMQMDRCADVLRQIRKTNMISAQSPFCIHLRLRSATAVSMVLAKA	3048
QY	89	-----PED-----	91
Db	3049	GGAFMPVDISQPRSRLONLIEESCAKLVLTLPESANALATLSGLTKVIPVLSLSELVQOIT	3108
QY	92	-----DTDP-----	95
Db	3109	DNTTKDEYCKSGDTDPSPAYLLYTSCTSGKPKGVVMEHRAWSLGTCHLSYMGFNSCT	3168
QY	96	-----	95
Db	3169	RILOFSSLMFDLSLEIWAIVYAGGCLFIPSDKERVNNLODFTRINDINTVLTPTSIGKL	3228
QY	96	-----	95
Db	3229	LNPXDLNPI SFAGIFGEMTRSLIDAMTLPGRRLVNSYGPTEACVLVTAREISPTAPHDK	3288
QY	96	-----GHA-----	98
Db	3289	PSSNIGHALGANIWWBPQRTALVPIGAVGELCTEAPSLARCYLEANPERTEYSPSTVLD	3348
QY	99	-----DLVLYIT-----	105
Db	3349	NWQTKGTRVYRTGDLVRYASDGTDLFLGRKDGQIKLRGQRIELGEIEHHIRRLMSDDPR	3408
QY	106	-----	105
Db	3409	FHEASVOLYNPATDPDRDATVDVQMRPEYLAGLLVLVLVSLMRSDSMHVLNIANTSENL	3468
QY	106	-----	105

```

Db 3469 QTLVTELKKSRLGVLPHYVPLHFVAVSLPTGSSGKLDHAFVRACLRELTAFLDGNFPK 3528
QY 106 -----
Db 3529 VEOVLTTNESVLQWNGTVLAMPDPSIQRGDDFFSLGGSSISAMRLVGLARSGHKLQHE 3588
QY 106 -----RPDL-----
Db 3589 DIFMCPRLADMAQISFVQESVPTTPTIKFDLLDDCEVIDHILPQLDMNKELIE 3648
QY 110 -----
Db 3649 DVYPTPLOSMAATARHGEAYTMQISITVLAQAQLKAMDVVFRDFEVLRTIALG 3708
QY 110 -----
Db 3709 PSQALQVVVKHELSWESFSPQSFKDHFYRSLGYGKPLARLAVITQALDKTQPSHGT 3768
QY 110 -----
Db 3769 REARTKNSQDTVMVVGAAHHSIYDAHVLMSIMRRLYREFITGSOADGILEAETSRSQGVVP 3828
QY 110 -----ELPD 113
Db 3829 FKSVEKLLRGKDNDESLFLFWKEXRGVSSQPPASPAPRVLEHQPSTATQTLTKVSLPT 3888
QY 114 GNRQVRGVT-----
Db 3889 SSRKLGATVATVAYAAWALTIHYTADPDVDFGATLSGRETNAGSISHPESITAGPTIIT 3948
QY 123 -----
Db 3949 VPLRIIDFQTVSDFLSTLQKDIVRAAYFGQMGLNSIAHIDNCDRCACGFKSIIVQV 4008
QY 123 -----
Db 4009 PDEGENHVGRAANPFQMSLESIGHFPAPLYVEVEQSESTDVLIRMAVDPLVPEKLAHFI 4068
QY 123 -----
Db 4069 SDTFTTMSNLNANPAKAKVESIPALSEAHLELDVTCPEWILGKAKDEKIRTESHQCLQ 4128
QY 123 -----QLG-----
Db 4129 DLVCRRAQSPNSQAIDSDMGDSISYHELDGLSSI LAEHLSQLGVPEAPVCLLFEKSKWA 4188
QY 126 -----
Db 4189 VVAMIGI KAGGCFVPLDPSYPHERLEHIISETGSSVITVSAAYSKLCLSLSVRGIVCDG 4248
QY 126 -----
Db 4249 SVFSSTKKPLPSTADSPSPSVRPNQAAIYLTFTSGTGKPKGVVMEHHSVCALIALGKR 4308
QY 126 -----GAC-----
Db 4309 MGLGQPSRVLPNSYFDMVLLDIFGLTVYGGCLCIPKESQMSNLGWSQKFKVNTMLL 4368
QY 129 -----
Db 4369 STSVSLRMQPADTPSLETCLTGEAVLQSDVDRWAPKHLIAGYGPETTCIMSVSGELTP 4428
QY 129 -----
Db 4429 SSPANLIGKPVSCQAWINPLKETELAPYATGELYTQGTPTVARGYLHDDVLTSKAFIVD 4488
QY 129 -----
Db 4489 PQWLTYKTNENQWSRRAYKTGDLVFWGPOSNLYYVRKDSQVKIRQORVELAEIEVI 4548
QY 129 -----
Db 4549 RQHIPPVTVCDLLSSDDQNTRIILGAVLIGDRALGGPEDLEVIGYMDLKSIIIPAL 4608

```

```

QY 129 -----SPT----- 131
Db 4609 EASLPHHMIPEAYVPFVQLPTLGSGKLDKRTVRRVAGFLAFSLPOASARHPNOPTVTHQ 4668
QY 132 -----
Db 4669 KULLRLMCKILPOLDESANVKNODNFGIGGDSIAAIIKLVALLRQHGISLAIAEITRPTL 4728
QY 132 -----
Db 4729 EAMSSLIIDEHNFVSHAGILSDVTRNTSGVMROTTNLIAGRHSMVAKESRECDNSTLPCT 4788
QY 132 -----
Db 4789 EYQOMFLAGTEAFTGAHSAQFIFRLPEKIDLRLOAAPHDCADWYPNLRTOIHKDADTOR 4848
QY 132 -----WSC-----
Db 4849 LLHDISPGVKVPWSCHYSDDLNTVLSHDKKFPGLDGLPHRVITMRHRDPTESMLVWTL 4908
QY 135 -----LITEDT-----
Db 4909 NHAAYDANSLRMWLEHITEAYANPDYPSYSLGWTAFLVLTENTKEASRSFWSVSLVDVK 4968
QY 141 -----
Db 4969 PARLMFNVLNVSINPRQDRLEARIINIKRVLSQLAATAATVLLAGLTLVARVCDTRDVILA 5028
QY 141 -----
Db 5029 HLLTGRTUPLAGIENCPGTTITKVPRLPLMDQDLVLELDSVAKKITHELMRMPHEHS 5088
QY 141 -----GFDL 144
Db 5089 GLSAIREFIPQAEGLTSSGKPHAGSVLGRPLDLVIHPKGLDL 5133

```

```

RESULT 6
BIR6 HUMAN
ID BIR6 HUMAN STANDARD; PRT: 4829 AA.
AC Q9NR09; Q9ULDI;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Baculoviral IAP repeat-containing protein 6 (Ubiquitin-conjugating
DE BIR-domain enzyme apollon).
GN BIRC6 OR KIAA1289.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20012759; PubMed=10544019;
RA Chen Z., Naito M., Hori S., Mashima T., Yamori T., Tsuruo T.;
RT "A human IAP-family gene, apollon, expressed in human brain cancer
RT cells.";
RL Biochem. Biophys. Res. Commun. 264:847-854(1999).
RN [2]
RP SEQUENCE OF 3238-4829 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RL DNA Res. 6:337-345(1999).
CC -!- FUNCTION: MAY PROTECT CELLS FROM UNDERGOING APOPTOSIS.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN CANCER CELLS.
CC -!- SIMILARITY: CONTAINS 1 BIR REPEAT.

```



```
Qy 66 ----- 65
Db 1641 VGPVHNSVPNPVAPAFIHPSDVIPPPTKPTPLFMTPLTPNEAVSVVINAELQLF 1700
Qy 66 ----- 72
Db 1701 PGSVIDPPAVNLAHNKNSKSRMNPGLSGLALAISHASHFLOPPPHQSIIRMHSGAR 1760
Qy 73 ----- 82
Db 1761 RFVTLDFGRPILLTDVLIPTCGDLASLSIDIWTLGEEVDGRRLVVATDSTHSLILHDLI 1820
Qy 83 ----- 83
Db 1821 PPPVCREMKITVIGRYGSTNARAKIPLGFFYGYHTYILPWESELKLMHDPKLGESANQP 1880
Qy 84 ----- 83
Db 1881 EIDQHLAMVALQEDIOCRYNLACHRLTLLQSIDLPLPLSANNAAQYFLRKPDKAVEEDS 1940
Qy 84 ----- 91
Db 1941 RVFSAYODCIQLQLNLNAHNAVORLKVAGASRKMLSETSNPEDLIQTSSTEOLRTIIR 2000
Qy 92 ----- 91
Db 2001 YLLDTLLSLLHASNHSVPVAVLQSTFHAQACEELFKHLCISGTPKIRLHTGLLLVLQCGG 2060
Qy 92 ----- 91
Db 2061 ERWMOFLSNVLOELYNSEQLIPPODRVFMLLSCIGQRSLSNGVLESLLNLDNLLSP 2120
Qy 92 ----- 91
Db 2121 LQQLPMHRRTEGVLDIPMISWVMLVSRLLDYVATVEDEAAAAKPPLNGQWSFINNL 2180
Qy 92 ----- 91
Db 2181 HTQSLNRSSKSSSLDRYSRKIRKQLVHHKQQLNLLKAKQKALVEQMEKEKIOSNGSS 2240
Qy 92 ----- 94
Db 2241 YKLLVEQAKLQATSKHFKDILRLRTAENSRSLNLTDEVTTAKESPEIEPLPTLAHERC 2300
Qy 95 ----- 106
Db 2301 ISVVQKLVFLFSMDFTCHADLLFLFVCKVLARIANATRPITHLCEIVNEPQLERLLLV 2360
Qy 107 ----- 106
Db 2361 GTDFNRCDISWGNWAQYSLTCMLQDILAGELLAPVAABEAGTVGDDYVATAGDSDDS 2420
Qy 107 ----- 106
Db 2421 LQSSVOLLETIDEPLTHDITGAPPLSLEKDEIDLELLQDLMVEDIDPLDIDLEKDL 2480
Qy 107 ----- 106
Db 2481 AAKVKPISSTWYDWGADYGTYNPNPYIGGLGIPVAKPPANTEKNGSQTVSVVSQALD 2540
Qy 107 ----- 106
Db 2541 ARLEVGLEQOALMLKMMSTLEADSILQALNTSPTLSQSPTGTDSDLGLQANQTSQ 2600
Qy 107 ----- 106
Db 2601 LIITQSSVPMNLVNCNFKLSMLQVHHVQLESLLQLWLTLSLNSSSGNKENGADIFLYNA 2660
Qy 107 ----- 106
Db 2661 NRIPVISLNOASITSFLTVLAWYNTLLRTWCLVLSLTLMTNQLNSGSSAIGTOEST 2720
```

```
Qy 107 ----- 106
Db 2721 AHLLVSDPNLIHVLVKFLSCTSPHGTNOHSPQVPTATOAMQEBFLTRLOVHLSSTCQPIF 2780
Qy 107 ----- 106
Db 2781 SEFLLKLIHILSTERGAFOTGQGPLDAQVKLEFTLEONPEVVSVTISAVIESVTFLVH 2840
Qy 107 ----- 109
Db 2841 HYITCSDKYMSRSGSDSSVGARACFGGLPANLIRPGDAKAVCGEMTRDQLMFDLLKLVNI 2900
Qy 110 ----- 115
Db 2901 LVQPLSGNREYGARVSVTNTTDSVDEBKSGDKGSGSSTVOGSPAYVADLVLANQ 2960
Qy 116 ----- 115
Db 2961 QIMSQILSALGLCNSAMAMIIGASGLHLTKHENFHGGLDAISVGDLFTILTTLTKKAS 3020
Qy 116 ----- 115
Db 3021 TVHMLQPILTVMACGYMGROGSLATCOLSEPLLWFLRVLDTSDALKAFHDMGGVQLIC 3080
Qy 116 ----- 115
Db 3081 NNMVTSTRAIVNTARSMVSTIMKFLDSGPNKAVDSTLKTILASEPDNAEGIHNFAPLGT 3140
Qy 116 ----- 119
Db 3141 ITSSPTAOPAEVLLQATPPHRRARSAAWSYIFLPBEAMCDLTIHLPAAVLLKEIHIQPH 3200
Qy 120 ----- 119
Db 3201 LASLATCPSSVSEVADGVNMLPLSTPVVTSGLTYIKIOLVKAEVASAVCLRHRPRDA 3260
Qy 120 ----- 119
Db 3261 STLGLSQIKLLGLTAFGTSSATVNNPFLPSEDQVSKTSIGWLRLLHCLTHISLEGMM 3320
Qy 120 ----- 119
Db 3321 ASAAAPTANLLOTCAALLMSPYCGMHSPNIEVVLVKIGLOSTRIGLKLIDILLRNCAASG 3380
Qy 120 ----- 123
Db 3381 SDPTDLNSPLLFRNLNGLSSDSTIDILYQLGTTQDPTKDRIOALLKWVSDSARVAAMKR 3440
Qy 124 ----- 123
Db 3441 SGRMYMCPNSSTVEYGLLMPSPSHLHVAAAILWHSYELLVEYDLPALLDQELFELLFNW 3500
Qy 124 ----- 123
Db 3501 SMSLPCNMVLKXAVDSLCSMCHVHPNYPFSLMGWNGITPPPVQCHHRLSMTDDSKKQDL 3560
Qy 124 ----- 123
Db 3561 SSSLTDDSKNAQAPLALTESHLATLASSSQSPPAIKQLLDGLPSLLVRSLSASFCSHIS 3620
Qy 124 ----- 128
Db 3621 SSESIAQSIDISQDKURRHHVPOCNKMPITADLVAPILRFLTEVGNHIMKDWLGSEV 3680
Qy 129 ----- 135
Db 3681 NPLWTALLFLCHSGSTSGSHNLGAOOTSARSASLSAATTLTTOORTAIENATVAFFL 3740
Qy 136 ----- 135
Db 3741 QCISCHPNKQKMAQVLCFQTSPOGNLPTSGNISGPIRRLFLQMLEDEKVTMFLOS 3800
Qy 136 ----- 135
```

DB 3801 PCPLYGRINATSHVIOHPMYGAGHKFRTLHPVSTTSLDVLDRVSDTPSITAKLISQK 3860
QY 136 ----- 135
DB 3861 DDKENKHEEKEKVAENGQDNYSVVASGLKSQKRAVSATPPRPPRRGRTIPDKIG 3920
QY 136 ----- 135
DB 3921 STSGAAANKIITVPFHLPHKLAGOPLPAEMTLAQLLTLLYDRKLPQGVRSIDLTVKL 3980
QY 136 ----- 135
DB 3981 GSRVITDPSLSTDSYKRLHPKDHGDLASCPEDALTPGDCMDGILDESLLTCTPIQ 4040
QY 136 ----- 135
DB 4041 SPLQVAGMGGLAIAERLPLVPEVIQVVSAPVVTSTTOEKPDKSDQPEWVTIEQSGL 4100
QY 136 ----- 135
DB 4101 VYEAPETVAEPPPIKSAVOTMSPAPHAFLRPLPGVAEVLKPKHQAQCLLRIV 4160
QY 136 --ITD----- 139
DB 4161 LGVTDGEGSHILOSPSANVPLTLPVHVSLSFTTTLTDDGVLLRRNALSIGALHLIL 4220
QY 140 -----TGFDLGV 147
DB 4221 VCLSALSHSPRVNPNVSSVNOTEPOVSSHNPTSTEBQQLYWAQGTGFGTST 4272
RESULT 7
FAT2 HUMAN STANDARD; PRT; 4349 AA.
AC Q9NYQ8: O75091; QNSR7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protocadherin Fat 2 precursor (hFat2) (Multiple epidermal growth
factor-like domains 1).
GN FAT2 OR MEGF1 OR CDHF8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20202599; PubMed=10716726;
RA Wu Q., Maniatis T.;
RT "Large exons encoding multiple ectodomains are a characteristic
feature of protocadherin genes.";
RT Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
RN [2]
RP SEQUENCE OF 3777-4349 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple
EGF-like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
RN [3]
RP SEQUENCE OF 4142-4349 FROM N.A.
RC TISSUE=Testis;
RA Pouetka A., Wellenreuther R., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 32 CADHERIN DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ G-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.eib.ch/announce/>
or send an email to license@eib.ch).
EMBL; AF231022; BAA32463.1; -
EMBL; AB011535; BAA32463.1; -
EMBL; AL157443; CAB75663.1; -
Genew; HGNC:3596; FAT2.
MIM; 604269; -
HSP; P00740; IEDM.
InterPro; IPR002126; Cadherin.
InterPro; IPR000561; EGF-like.
Pfam; PF00028; cadherin; 33.
Pfam; PF00008; EGF; 2.
PRINTS; PRO0205; CADHERIN.
SMART; SM00112; CA; 32.
SMART; SM00181; EGF; 2.
SMART; SM00282; LamG; 1.
PROSITE; PS00232; CADHERIN_1; 14.
PROSITE; PS00268; CADHERIN_2; 32.
PROSITE; PS00022; EGF_1; 2.
PROSITE; PS01186; EGF_2; 1.
PROSITE; PS00025; LAM_G_DOMAIN; 1.
Transmembrane; Glycoprotein; Repeat; EGF-like domain; Signal.
CHAIN 1 18
DOMAIN 19 4349 PROTOCADHERIN FAT 2.
TRANSMEM 4049 4069 EXTRACELLULAR (POTENTIAL).
DOMAIN 4070 4349 CYTOPLASMIC (POTENTIAL).
DOMAIN 34 148 CADHERIN 1.
DOMAIN 149 256 CADHERIN 2.
DOMAIN 363 458 CADHERIN 3.
DOMAIN 459 564 CADHERIN 4.
DOMAIN 565 669 CADHERIN 5.
DOMAIN 716 820 CADHERIN 6.
DOMAIN 821 925 CADHERIN 7.
DOMAIN 926 1032 CADHERIN 8.
DOMAIN 1033 1137 CADHERIN 9.
DOMAIN 1138 1242 CADHERIN 10.
DOMAIN 1243 1346 CADHERIN 11.
DOMAIN 1350 1448 CADHERIN 12.
DOMAIN 1449 1555 CADHERIN 13.
DOMAIN 1556 1660 CADHERIN 14.
DOMAIN 1661 1758 CADHERIN 15.
DOMAIN 1759 1872 CADHERIN 16.
DOMAIN 1969 2070 CADHERIN 17.
DOMAIN 2071 2171 CADHERIN 18.
DOMAIN 2172 2272 CADHERIN 19.
DOMAIN 2273 2379 CADHERIN 20.
DOMAIN 2380 2481 CADHERIN 21.
DOMAIN 2482 2585 CADHERIN 22.
DOMAIN 2586 2691 CADHERIN 23.
DOMAIN 2692 2797 CADHERIN 24.
DOMAIN 2798 2906 CADHERIN 25.
DOMAIN 2907 3011 CADHERIN 26.
DOMAIN 3012 3113 CADHERIN 27.
DOMAIN 3114 3218 CADHERIN 28.
DOMAIN 3219 3321 CADHERIN 29.
DOMAIN 3322 3426 CADHERIN 30.
DOMAIN 3427 3531 CADHERIN 31.
DOMAIN 3532 3642 CADHERIN 32.
DOMAIN 3643 3744 LAMININ G-LIKE.
DOMAIN 3745 3844 EGF-LIKE 1.
DOMAIN 3845 3949 EGF-LIKE 2.
DOMAIN 3950 4022 POLY-SER.
DOMAIN 4023 4102 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 4103 4200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 4201 4300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 4301 4400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 4401 4500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 4501 4600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 4601 4700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 4701 4800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 4801 4900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 4901 5000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 5001 5100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 5101 5200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 5201 5300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 5301 5400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 5401 5500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 5501 5600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 5601 5700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 5701 5800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 5801 5900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 5901 6000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 6001 6100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 6101 6200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 6201 6300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 6301 6400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 6401 6500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 6501 6600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 6601 6700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 6701 6800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 6801 6900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 6901 7000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 7001 7100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 7101 7200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 7201 7300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 7301 7400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 7401 7500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 7501 7600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 7601 7700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 7701 7800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 7801 7900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 7901 8000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 8001 8100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 8101 8200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 8201 8300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 8301 8400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 8401 8500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 8501 8600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 8601 8700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 8701 8800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 8801 8900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 8901 9000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 9001 9100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 9101 9200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 9201 9300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 9301 9400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 9401 9500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 9501 9600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 9601 9700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 9701 9800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 9801 9900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 9901 10000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 10001 10100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 10101 10200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 10201 10300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 10301 10400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 10401 10500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 10501 10600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 10601 10700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 10701 10800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 10801 10900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 10901 11000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 11001 11100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 11101 11200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 11201 11300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 11301 11400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 11401 11500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 11501 11600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 11601 11700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 11701 11800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 11801 11900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 11901 12000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 12001 12100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 12101 12200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 12201 12300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 12301 12400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 12401 12500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 12501 12600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 12601 12700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 12701 12800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 12801 12900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 12901 13000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 13001 13100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 13101 13200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 13201 13300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 13301 13400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 13401 13500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 13501 13600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 13601 13700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 13701 13800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 13801 13900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 13901 14000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 14001 14100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 14101 14200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 14201 14300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 14301 14400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 14401 14500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 14501 14600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 14601 14700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 14701 14800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 14801 14900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 14901 15000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 15001 15100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 15101 15200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 15201 15300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 15301 15400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 15401 15500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 15501 15600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 15601 15700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 15701 15800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 15801 15900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 15901 16000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 16001 16100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 16101 16200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 16201 16300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 16301 16400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 16401 16500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 16501 16600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 16601 16700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 16701 16800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 16801 16900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 16901 17000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 17001 17100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 17101 17200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 17201 17300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 17301 17400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 17401 17500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 17501 17600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 17601 17700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 17701 17800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 17801 17900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 17901 18000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 18001 18100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 18101 18200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 18201 18300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 18301 18400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 18401 18500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 18501 18600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 18601 18700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 18701 18800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 18801 18900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 18901 19000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 19001 19100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 19101 19200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 19201 19300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 19301 19400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 19401 19500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 19501 19600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 19601 19700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 19701 19800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 19801 19900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 19901 20000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 20001 20100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 20101 20200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 20201 20300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 20301 20400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 20401 20500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 20501 20600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 20601 20700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 20701 20800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 20801 20900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 20901 21000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 21001 21100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 21101 21200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 21201 21300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 21301 21400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 21401 21500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 21501 21600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 21601 21700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 21701 21800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 21801 21900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 21901 22000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 22001 22100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 22101 22200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 22201 22300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 22301 22400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 22401 22500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 22501 22600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 22601 22700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 22701 22800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 22801 22900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 22901 23000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 23001 23100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 23101 23200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 23201 23300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 23301 23400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 23401 23500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 23501 23600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 23601 23700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 23701 23800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 23801 23900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 23901 24000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 24001 24100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 24101 24200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 24201 24300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 24301 24400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 24401 24500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 24501 24600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 24601 24700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 24701 24800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 24801 24900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 24901 25000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 25001 25100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 25101 25200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 25201 25300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 25301 25400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 25401 25500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 25501 25600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 25601 25700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 25701 25800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 25801 25900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 25901 26000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 26001 26100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 26101 26200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 26201 26300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 26301 26400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 26401 26500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 26501 26600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 26601 26700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 26701 26800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 26801 26900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 26901 27000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 27001 27100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 27101 27200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 27201 27300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 27301 27400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 27401 27500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 27501 27600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 27601 27700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 27701 27800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 27801 27900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 27901 28000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 28001 28100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 28101 28200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 28201 28300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 28301 28400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 28401 28500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 28501 28600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 28601 28700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 28701 28800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 28801 28900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 28901 29000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 29001 29100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 29101 29200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 29201 29300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 29301 29400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 29401 29500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 29501 29600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 29601 29700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 29701 29800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 29801 29900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 29901 30000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 30001 30100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 30101 30200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 30201 30300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 30301 30400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 30401 30500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 30501 30600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 30601 30700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 30701 30800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 30801 30900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 30901 31000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 31001 31100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 31101 31200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 31201 31300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 31301 31400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 31401 31500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 31501 31600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 31601 31700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 31701 31800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 31801 31900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 31901 32000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 32001 32100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 32101 32200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 32201 32300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 32301 32400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 32401 32500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 32501 32600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 32601 32700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 32701 32800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 32801 32900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 32901 33000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 33001 33100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 33101 33200 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 33201 33300 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 33301 33400 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 33401 33500 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 33501 33600 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 33601 33700 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 33701 33800 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 33801 33900 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 33901 34000 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 34001 34100 N-LINKED (GLCNAC. . .) (POTENTIAL).
DOMAIN 34101 34200 N-LINKED (GLCNAC

FT	CARBOHYD	627	627	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	655	655	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	789	789	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	996	996	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	1175	1175	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	1303	1303	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	1383	1383	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	1417	1417	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	1904	1904	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	1998	1998	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	2007	2007	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	2165	2165	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	2183	2183	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	2325	2325	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	2368	2368	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	2387	2387	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	2430	2430	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	2470	2470	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	2547	2547	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	2597	2597	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	2654	2654	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	3125	3125	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	3276	3276	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	3310	3310	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	3430	3430	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	3471	3471	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	3601	3601	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	3772	3772	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	3813	3813	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	3840	3840	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	3873	3873	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	3904	3904	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CARBOHYD	3989	3989	N-LINKED	(GLCNAC. .)	(POTENTIAL)
FT	CONFLICT	4117	4117	P -> L (IN REF. 2).		
FT	CONFLICT	4160	4160	E -> G (IN REF. 2).		
SQ	SEQUENCE	4349	4349	AA; 479383 MW; 6C050CA17AD398ED CRC64;		
Query Match 27.8%; Score 215; DB 1; Length 4349;						
Best Local Similarity 2.0%; Pred.No. 24;						
Matches 88; Conservative 25; Mismatches 24; Indels 4198; Gaps 29;						
Qy	5	ILH-----	7			
Db	12	LHCAECXPLEGILSSAHFTHSHYNATYENSSPKTYVESFERQGIYLAEPQWAVRY 71				
Qy	8	-----LEL----- 10				
Db	72	RIISGDVANFKTEYVVGNFCLRTKSSNTALLNREVRSYTLIIQATEKTELEAL 131				
Qy	11	----- 10				
Db	132	TRVVHILDONLUKPLFSPSYRVYTISEDMPKSPICKVTATDADLGQNAEFYAFNTRS 191				
Qy	11	----- 10				
Db	192	EMFAIHTSGVTVVAGKLVNTHRGKHELQVLAVDRMRKI SEGNGFGSLAALVVHVEPALR 251				
Qy	11	----- 10				
Db	252	KPPAIASVVVTPPDSNDGTTYATVLVDANSAGAEVSEVVGDPGPKFKAIKSYARSNE 311				
Qy	11	--LVAV----- 14				
Db	312	FSLSVKNDINMWEYLHGNLSLOARSGSGPYFYQIRGFHLPPSKLSSLKFEKAVRVQL 371				
Qy	15	----- 14				
Db	372	SEFSPGSRVVMVTPAFNLOVVLKPPSENVGFKLNARTGLITTKLMDFDHRAHYQL 431				
Qy	15	----- 14				
Db	432	HIRTSPQASTVVVIDIVDCNNHAPLNRSSYDGTLDENIPPGTSVLAVTATDRDHGENG 491				

Qy	15	-----GP----- 16				
Db	492	YVTVSIAGPKALPFSIDPYLGIISTSKPMDELMKRIYTFRVRASDNGSPRREKEVSIF 551				
Qy	17	-----DV----- 18				
Db	552	LQRLNLDNOMPFEEVNCSTSIRODWPVGKSIIMTMSAIDVDELQNLKYEIVSGNELEYFD 611				
Qy	19	----- 18				
Db	612	LNHFSGVISLKRPFINLTAGQPTSYSCLKITASDCKNVASPTTLNITVVVKDPHFPEVPTCD 671				
Qy	19	-----FOAH----- 22				
Db	672	KTGVLTOFTKILHFIQNGOESSDEFTSLSTYQINHYTPOFEDHPQS:DVLESVPIN 731				
Qy	23	----- 22				
Db	732	TPLARLAATDPDAGFNCKLVVVIADGNEEGCFDIELETGLTVAAPLDYEATNFVILNVT 791				
Qy	23	-----QEDTE----- 27				
Db	792	VYDLGTPOKSSMKLLTVNVKMDNDNAPRFPGGYQLTISEDTEVGTITIAELTTKDAISED 851				
Qy	28	----- 27				
Db	852	NGRVRYTLLSPTEKFSHLPLTGELVVTGHLDRSEPRYLKVEARDOPSKGHOLFVSVDL 911				
Qy	28	-----RYVL----- 31				
Db	912	ITLEDVNDNSPOCITEHNRUKLPEDLPPGTVLTFDASDPDLGPAGEVRVVLMDGANGT 971				
Qy	32	----- 31				
Db	972	FRVDMTGALLERBELDEFERBAGYNLSLWASDGRPLARRTLCHVEVIVLDVNNENLHPPH 1031				
Qy	32	----- 31				
Db	1032	FASFVHQGOVQENSPSGTQVIWAAQDDSGLDGLOVFLRAGTGLAAFSINQDTGMIQT 1091				
Qy	32	----- 31				
Db	1092	LAPLDREFASYWLTVLAVDRGVSPLSSVTEVYIEVTANDNPPQMSOAVFYPSIOEDAP 1151				
Qy	32	----- 31				
Db	1152	VCTSVLQLDANDPSSSKGKLTFNITSGNYMGFFMIHPVTGLLSTAQOLDRENKDEHILE 1211				
Qy	32	----- 31				
Db	1212	VTVLONGEPLSKSTRVVVGILDVNDNPPIFSHKLFNVRLPERLSPSPGPVYRLVASDL 1271				
Qy	32	----- 31				
Db	1272	DEGLNGRVTSIEDSYBEAFSIDLVTVGVSSNSTFTAGEYNILTIKATDSGOPPLSASVR 1331				
Qy	32	----- 31				
Db	1332	LHIEWIPMRPSSIPLAFDETYYSFTVMTDPVNHVGVISVEGRPLFWFNI SGGDKDM 1391				
Qy	32	-----TNLNI----- 36				
Db	1392	DFDIEKTTGSIVIARPLDTRRRSNYNLTVEVTDGSRITATQVHIFMIANINHHRPQFLET 1451				
Qy	37	-----GAECLR----- 42				
Db	1452	RYEVRVPODTVEGVELLRVAIDQDKGSLIYTHIGSQDPGSASLFOLDPPSGSVLTVCK 1511				
Qy	43	----- 42				
Db	1512	LDLGSGPSOHTLTVMRDQEIPIKRFVNVTHVEDGNLHPPRFTQLHYEASVPDTIAPG 1571				
Qy	43	----- 42				

```
Db 1572 TELLQVRAMDADRGVNAVEHYSLKGNSEGFNINALLGIITLAQKLDQANHAPHTLTVK 1631
Qy 43 ----- 42
Db 1632 AEOGSPQWHDLATVIIHVYSDRSAPIFSKSEYFVEIPESIPVGSPIILLVSAMSPSEVT 1691
Qy 43 ----- 42
Db 1692 YELREGNKQGVFSMNSYSGLIISTQKLDHEKISSYQKIRGSNNAGAFDTMVVVVDIIDE 1751
Qy 43 ----- 42
Db 1752 NDNAPMLFKSTFVQIIEAAPLYSMINDKXNNPFVHASDSDKEANSLLYVKILEPEALK 1811
Qy 43 ----- 53
Db 1812 FFKIDPSMGTLTIVSEMDYESMPSFQFCVYVHDQGSPLFAPRAQVIIHVRDNDSPPR 1871
Qy 54 ----- 53
Db 1872 FSEQIYEVAIVGPIHPGMELLMVRADEDESEVNSYIKTGNADSAVTIHPVTGSIISVLNPA 1931
Qy 54 ----- 60
Db 1932 FLGLSRKLTIRASDGLYQDTALVKISLTQVLDKSLQFDQDVYAAVKENLQDRKALVILG 1991
Qy 61 ----- 60
Db 1992 AQGNHLNDTLYFLNGTDMFHMVQSAQVLQTRGVAFDREQQDTHELAVEVRDNRTPQV 2051
Qy 61 ----- 63
Db 2052 AQGLVRVSIEDVNDNPPKFKLPHYTYIIQDGTBPGDVLQVQSATDEDLGTNGAVTYEFAE 2111
Qy 64 ----- 63
Db 2112 DYTFRIDPYLDISLKKPFQYQALNKYHLKVARDGGTSLQSEEEVLTVRNKSNPLF 2171
Qy 64 ----- 63
Db 2172 QSPYKVRVPENITLYTPILHTQARSPEGLRIYNIIVEEPMLFTTDFKTGVLTVTGPL 2231
Qy 64 ----- 67
Db 2232 DYESTKRVFTVRATDTALGSFSEATVEVLVEDVNDNPPTFPSQLVYTTSISEGLPAQTPV 2291
Qy 68 ----- 67
Db 2292 IQLLASQDSGRNRDVSQIVEDGSDSVSKPFQINGSTGEMSTVQELDYEAQQHPHVKVRA 2351
Qy 68 ----- 83
Db 2352 MDKGDPLTGELTVVNVVSDINDNPPPEFRPOVEANV--SELATCGHLVLKVQAIDPDSR 2409
Qy 84 ----- 83
Db 2410 DTSRLEYLILSGNODRHHFINSSGIIIMFNLCCKHLDSSYNLRVGASDGVFRATVPYI 2469
Qy 84 ----- 83
Db 2470 NTTNANKYSPEFQOQLYEAELAENAMVGTKVIDLALTDKDSGPGYGTIDIYTIINKLASEKF 2529
Qy 84 ----- 83
Db 2530 SINPGQIATLQKLDRENSTERVIAIKVMARDGGGRVAFCTVKIILTDENDNPPQPKASE 2589
Qy 84 ----- 91
Db 2590 YTVSIOQNVSKDSPVIOVLAYDADEGQNAVDTYSVNPEDLVKDVIEINPVTGVVVKKDSL 2649
Qy 92 ----- 91
```

```
Db 2650 VGLENQTLDFFIKAQDGGPPHMNSLVPVRLQVVPKKSLPFSFSEPLYTFSAPEDLPEGSE 2709
Qy 92 ----- 91
Db 2710 IGIVKAVAAQDPVIYSLVRGTTPESENKOGVFSLOPDGTGIVKVRKPMDESTKLYQIDVMA 2769
Qy 92 ----- 91
Db 2770 HCLQNTDVSLSVNIQVGDVNDNRVPVFEADPYKAVLTENMPVGTSVIOVTAIDKOTGRD 2829
Qy 92 ----- 91
Db 2830 GOVSRLSADPGSNVHELFAIDSESGWITTLQELDCETCQTYHFHVAVYDHGQTIQLSSQ 2889
Qy 92 ----- 91
Db 2890 ALVQVSIITDENDNAPRFASEEYRGSVVENSEPGEVLATUKTLADADISEONRQVTCVITEG 2949
Qy 92 ----- 91
Db 2950 DPLQGFGISOVDGDEMRISSRKTLDREHTAKYLLRVTASDGKFQASVTVEIFVLVDVNDNSP 3009
Qy 92 ----- 97
Db 3010 QCSQLLYTGKMHEDVPPGHFILKVSATDLDTDNAQIITYSLHGPGAHEFKLDPHTGELTT 3069
Qy 98 ----- 104
Db 3070 LTALDRERKDVNLVAKATDGGGRSQADITLHVEDVNDNAPRFPSPSHCAVAFDNTTVK 3129
Qy 105 ----- 104
Db 3130 TPVAVVPARDPQGANAOVVYSLPDSAEGHFSIDATTGVIRLEKPLQVRPOAPLELTVRA 3189
Qy 105 ----- 113
Db 3190 SDLGTBPLSTLGTVTVSVVGLDYLPVFLNTEHSVQVPEDAPPGTEVLQLATLTRPGAE 3249
Qy 114 ----- 119
Db 3250 KTGVRVYSGNEQGRFLDARTGILYVNASLDPETSPKYFLSIECSRKSSSSLSDDVTVMV 3309
Qy 120 ----- 119
Db 3310 NITDVNEHRPQPDQPYSTRVLENALVGDVILTVSATDEGPLNSDITYSLIGGNQLGHF 3369
Qy 120 ----- 119
Db 3370 TIHPKKGELQAKALDREQASSYSLKLRTATDSQPPLHEDTDIAIQVADVNDNPPRFFQL 3429
Qy 120 ----- 119
Db 3430 NYSTTVQENSPIGSKVLQILSDPSDPENGPYPSPFRITKGNNGSAPRVTPDGLVLTAEGL 3489
Qy 120 ----- 119
Db 3490 SRRAEWYQLOIQASDGPPLSSLTSVRVHVTEQSHYAPSLPLBIFITVGEDEFQGM 3549
Qy 120 ----- 119
Db 3550 VGKIHATDRDPQDTLTYSLAEBETLGRHFSVGAPDGKIIAAOGLPRGHYSFNVTVSDGTF 3609
Qy 120 ----- 119
Db 3610 TTTAGVHVYVHVHQBALQOAMMGFYQLTPEELVSDHWRNLQRFSLKLDIKRANIHLA 3669
Qy 120 ----- 119
Db 3670 SLOPAEAVAGVDVLLVFECHSGTFFEFQELASIIITHSAKEMEHSVGVQMRSAMPMVPCOG 3729
Qy 120 ----- 119
Db 3730 PTCQGOICHNTVHLDPKVGPTYSTARLSILTPRHHLQRCSCNGTATRFSGQSVRYRAP 3789
```

```
OY 120 -----GVTQL----- 124
Db 3790 AARNWHIHYLKTLPQAIILFTNETASVSLKASGVQLEHYHCLGFFYGNLSSQRHVND 3849
OY 125 ----- 124
Db 3850 HEWHSILVEEDASIRLMVDSMONTSLAVPENCRLRPERHLLGLILLHSSSNVSOGF 3909
OY 125 -----GGACSPWTS-- 133
Db 3910 EGCLDAVVNEEALDLLAPGKTVAGLLETQALTOCCLHSDYCSQNTCLNGGKCS--WTHG 3967
OY 134 ----- 133
Db 3968 AGYVCKPQPFQSKHCEQRENCIFAPCLEGGTCILSPKASGNCNCPHYTGDRCEMARG 4027
OY 134 -----CLIT----- 137
Db 4028 CSEGHCLVTEIQRDWGOQELLITVAVAFIITVGLLFYCRCKSHKPVAMEDPDLL 4087
OY 138 ----- 137
Db 4088 ARSVGVDTOAMPAIELNPLSASNNLNQPEPSKASVPNELVTFGPNSKORPVVCSVPR 4147
OY 138 ----- 137
Db 4148 LPPAAVPSHSDNEPVIKRTWSSEEMVYPGAMVWPPTYSRNERWEYPHSEVTQGPLPSPA 4207
OY 138 ----- 137
Db 4208 HHRSTPVMPEPNGLYGGFPFPLEMENKRAPLPPRYSNQNLDMPSRPRRLVAPC 4267
OY 138 -----EDTGF----- 142
Db 4268 LNEYTAISYHHSQFROGGGCPCLADGGYKGVGMRLSRAGPSYAVCEVAGLPGQCPQVR 4327
OY 143 -----DLG 145
Db 4328 PPNYEGSDMVESDYG 4342

RESULT 8
ID RYR1_PIG STANDARD; PRT; 5035 AA.
AC P16360;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ryanodine receptor 1 (Skeletal muscle-type ryanodine receptor) (RYR1)
DE (RYR-1) (Skeletal muscle calcium release channel).
GN RYR1 OR CRC.
OS Sus acrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Norwegian Landrace; TISSUE=Skeletal muscle;
RX MEDLINE=93036581; PubMed=1329581;
RA Harbitz I., Kristensen T., Bosnes M., Kran S., Davies W.;
RT "DNA sequence of the skeletal muscle calcium release channel cDNA and
RT verification of the Arg615-->Cys615 mutation, associated with porcine
RT malignant hyperthermia, in Norwegian landrace pigs.";
RL Anim. Genet. 23:395-402 (1992).
RN [2]
RP SEQUENCE OF 1129-2801 FROM N.A.
RN Brenig B.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1129-2643 FROM N.A.
RC STRAIN=German Landrace; TISSUE=Liver;
RX MEDLINE=94117003; PubMed=8288238;
```

```
RA Leeb T., Schmolzl S., Brem G., Brenig B.;
RT "Genomic organization of the porcine skeletal muscle ryanodine
RT receptor (RYR1) gene coding region 4624 to 7929.";
RL Genomics 18:349-354 (1993).
RN [4]
RP SEQUENCE OF 4785-5035 FROM N.A.
RX MEDLINE=91065640; PubMed=2174405;
RA Harbitz I., Chowdhary B., Thomsen P.D., Davies W., Kaufman U.,
RA Kran S., Gustavsen I., Christensen K., Hauge J.G.;
RT "Assignment of the porcine calcium release channel gene, a candidate
RT for the malignant hyperthermia locus, to the 6p11-->q21 segment of
RT chromosome 6.";
RL Genomics 8:243-248 (1990).
CC -!- FUNCTION: COMMUNICATION BETWEEN TRANSVERSE-TUBULES AND
CC SARCOPLASMIC RETICULUM. CONTRACTION OF SKELETAL MUSCLE IS
CC TRIGGERED BY RELEASE OF CALCIUM IONS FROM SR FOLLOWING
CC DEPOLARIZATION OF T-TUBULES.
CC -!- SUBUNIT: HOMOTETRAMER (POTENTIAL).
CC -!- MISCELLANEOUS: The calcium release channel is modulated by calcium
CC ions, magnesium ions, ATP and calmodulin.
CC -!- MISCELLANEOUS: The calcium release channel activity resides in the
CC C-terminal region while the remaining part of the protein
CC constitutes the 'foot' structure spanning the junctional gap
CC between the SR and the T-tubule. It is possible that the foot
CC structure interacts with the cytoplasmic region of the
CC dihydropyridine receptor.
CC -!- MISCELLANEOUS: Ryanodine is an alkaloid that binds to the Ca-
CC release channel in junctional SR and modulates its activity.
CC -!- SIMILARITY: BELONGS TO THE RYANODINE RECEPTOR FAMILY.
CC -!- SIMILARITY: CONTAINS 3 SPRY DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X62880; CAA44674.1; ALT_SEQ.
DR EMBL; X68247; CAA48318.1; -.
DR EMBL; X69465; CAA49225.1; -.
DR EMBL; M32501; AAA31022.1; -.
DR PIR; A37105; A37105.
DR InterPro; IPR000699; Ca-rel channel.
DR InterPro; IPR001682; Ca/Na_pore.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000636; M-channel_nlg.
DR InterPro; IPR003608; MIR.
DR InterPro; IPR003032; RYR.
DR InterPro; IPR001215; Ryanodn receptor.
DR InterPro; IPR003878; SPRY_domain.
DR InterPro; IPR003877; SPRY_receptor.
DR Pfam; PF00520; Ion_trans; 1.
DR Pfam; PF01365; RYDR_ITPR; 3.
DR Pfam; PF02026; RYR; 4.
DR Pfam; PF02815; MIR; 4.
DR PRINTS; PR00795; RYANODINER.
DR SMART; SM00472; MIR; 4.
DR SMART; SM00449; SPRY; 3.
KW Receptor; Transmembrane; Ionic channel; Calcium channel; Repeat;
KW Phosphorylation; Glycoprotein.
FT DOMAIN 1 3124
FT TRANSMEM 3125 3145 M' (POTENTIAL).
FT TRANSMEM 3189 3207 M' (POTENTIAL).
FT TRANSMEM 3980 3999 M1 (POTENTIAL).
FT TRANSMEM 4018 4036 M2 (POTENTIAL).
FT TRANSMEM 4274 4297 M3 (POTENTIAL).
FT TRANSMEM 4339 4359 M4 (POTENTIAL).
FT TRANSMEM 4557 4578 M5 (POTENTIAL).
FT TRANSMEM 4646 4669 M6 (POTENTIAL).
FT TRANSMEM 4787 4807 M7 (POTENTIAL).
```

FT	TRANSMEM	4835	4854	M8 (POTENTIAL).	Db	1156	TENTIIFTLNGEVLMSDGSSETAFRDIEVGDGPLVCSLPGQVGHNLNGDVSSLRFFA	1215
FT	TRANSMEM	4877	4896	M9 (POTENTIAL).	Qy	45	-----	44
FT	TRANSMEM	4921	4935	M10 (POTENTIAL).	Db	1216	ICGLQGEFEPFAINMORPVTTWFSKSLPQFEAVPLEHPHYEVSVDGTVDTPPCRLRTHR	1275
FT	DOMAIN	659	797	SPRY 1.	Qy	45	-----	54
FT	DOMAIN	1085	1208	SPRY 2.	Db	1276	TWGSQNSLVEMFLRLSLPVQFHQFRCTAGATPLAPGLOPPAEDEARAAEPDPDYENL	1335
FT	DOMAIN	1430	1570	SPRY 3.	Qy	55	-----	VK 56
FT	DOMAIN	1474	1525	GLU-RICH (ACIDIC).	Db	1336	RRSAGRNGEAGGKEGTAKGAPGTAQAGVEAPPAENENEKDATTEKNKKRGFLFAKK	1395
FT	DOMAIN	1874	1925	PRO-RICH.	Qy	57	MVILTEPEGAPNI	69
FT	DOMAIN	4458	4526	6 X APPROXIMATE REPEATS.	Db	1396	AAMTQPPATPLRLPHEVVPADDDPDIILNTTYYYSRVFAGOEPCVMVGWTP	1455
FT	DOMAIN	842	2360	1.	Qy	70	-----	69
FT	REPEAT	842	955	2.	Db	1456	DYHQHDMNFDLTQVRAVTVTMGDEQNIHSLKSCNCVMWGGDFVSPGOQGRISHTDLV	1515
FT	REPEAT	956	1069	3. (INCOMPLETE).	Qy	70	-----	72
FT	REPEAT	1345	1360	4. (INCOMPLETE).	Db	1516	IGCLVDLATGLMTFTANGKESNTFFQVEPNTKLPFAVFLPTHONVIOFELGKQKNIMPL	1575
FT	REPEAT	1373	1388	5.	Qy	73	-----	72
FT	REPEAT	2727	2846	6.	Db	1576	SAAMFLSERKNPAPQCPRLPMLQMPVSWSRMPNHLFVETRRAGERLGVAVCOEPLT	1635
FT	REPEAT	2847	2860	PHOSPHORYLATION (BY PKA AND PKG) (BY	Qy	73	-----	72
FT	MOD_RES	2844	2844	SIMILARITY).	Db	1636	MMALHIPEENRCMDILSERLDLQFHSHTLRLYRAVCAKNNRVAHALCSHVDAQALL	1695
FT	MOD_RES	3947	3947	PHOSPHORYLATION (POTENTIAL).	Qy	73	-----	72
FT	MOD_RES	4320	4320	PHOSPHORYLATION (POTENTIAL).	Db	1696	HALEDAHLPGPLRAGYVDLLSIHLESACRSRRSMLSEYIVPLTPETRAITLFPFGKRT	1755
FT	CARBOHYD	3467	3467	N-LINKED (GLCNAC. . .) (POTENTIAL).	Qy	73	-----	76
FT	CARBOHYD	3475	3475	N-LINKED (GLCNAC. . .) (POTENTIAL).	Db	1756	NGPRRHGLPGVGVTTLSRPPHFSAPCFVAALPAVGAAEAPARLSPSIPLEBALRDKALRM	1815
FT	CARBOHYD	3304	3304	N-LINKED (GLCNAC. . .) (POTENTIAL).	Qy	77	-----	76
FT	CARBOHYD	3945	3945	N-LINKED (GLCNAC. . .) (POTENTIAL).	Db	1816	LGEAVROGGQHARDPVGSGVEFQFPVVLKLVSTLLVMGIFGDEVDVKOILKMIIEVPTTE	1875
FT	CARBOHYD	4144	4144	N-LINKED (GLCNAC. . .) (POTENTIAL).	Qy	77	-----	76
FT	CARBOHYD	4862	4862	N-LINKED (GLCNAC. . .) (POTENTIAL).	Db	1876	EE	1935
FT	CARBOHYD	4862	4862	N-LINKED (GLCNAC. . .) (POTENTIAL).	Qy	77	-----	76
FT	CONFLICT	2092	2092	A -> P (IN REF. 2 AND 3).	Db	1936	VKLOWCNLLFYCDOELQHRVESLAFAERYVDKLOANORDRYGILMKAFMTAETARR	1995
FT	SEQUENCE	5035	5035	AA; E00613F2027B94A4 CRC64;	Qy	77	-----	82
Query Match 27.4%; Score 212; DB 1; Length 5035;								
Best Local Similarity 2.6%; Pred. No. 51;								
Matches 83; Conservative 32; Mismatches 33; Indels 3096; Gaps 27;								
Qy	1	AAGGIHL	-----	8	Db	1636	MMALHIPEENRCMDILSERLDLQFHSHTLRLYRAVCAKNNRVAHALCSHVDAQALL	1695
Db	556	ASSGILEVLYCVLIESPEVNIQENHIKSIISLLDKHGRNHKVLVDLCLVCNGVAVR	615		Qy	73	-----	72
Qy	9	-----ELLV-----AVGPDVFOAHOEDTERY-----	29		Db	1696	HALEDAHLPGPLRAGYVDLLSIHLESACRSRRSMLSEYIVPLTPETRAITLFPFGKRT	1755
Db	616	SNQDLITENLLPGRELLQTLNLYNVTISRPNIFVGRAGETTOYKWKYFEVWVDEVVFL	675		Qy	73	-----	76
Qy	30	-----VLT-----	32		Db	1756	NGPRRHGLPGVGVTTLSRPPHFSAPCFVAALPAVGAAEAPARLSPSIPLEBALRDKALRM	1815
Db	676	TAQATHLRVGMALTEGYSYPGCGEGGNGVGDLLYSYFGDGLHWTGHVPRLVTSPOQ	735		Qy	77	-----	76
Qy	33	-----NLN-----	35		Db	1816	LGEAVROGGQHARDPVGSGVEFQFPVVLKLVSTLLVMGIFGDEVDVKOILKMIIEVPTTE	1875
Db	736	HLLAPEDVVSCCLDLSPSISFRINGCPVQGVFEAFNLNGLFFPVVFSAGVKVFRLLGG	795		Qy	77	-----	76
Qy	36	-----	35		Db	1876	EE	1935
Db	796	RHGEFKFLPPPGYAPCHEAVLPRERLRLPEIKYRREGPRGRPHLVGPSRCLSHTDVPCP	855		Qy	77	-----	76
Qy	36	-----	35		Db	1936	VKLOWCNLLFYCDOELQHRVESLAFAERYVDKLOANORDRYGILMKAFMTAETARR	1995
Db	856	VDTQVILPPLHLIRERKLAENIHELWALTRIEQGWYGPVRDDNKRHLPCLDVDFHSUPE	915		Qy	77	-----	82
Qy	36	-----	35		Db	1996	TREPRSPQEOINMLLHPKQGEDEEDCPLPDEIRODLLFHFQDILLTHCGIOLEGESEEP	2055
Db	916	PERNVNLMQSGETLKTLLALGCHVMADEKADENLRKTKLPKTYMMSNGYKAPLDLSHV	975		Qy	83	-----	85
Qy	36	-----	35		Db	2056	EEATLGSRLMSLEKVLVKKEKSEBEPPEAESKAQSLQELVSHTVVRQAQEDFVQSP	2115
Db	976	RLTPAQTTLVRLAENGHNWADRVAQGSYSVAVQDIPARRNPLRVLYRLLDEATKRN	1035		Qy	86	-----	93
Qy	36	-----	35		Db	2116	ELVRAMFSLHRQVDGLGELLRALPRAYTISPSVDEMTSLLECIGQIRSLIIVOMQOE	2175
Db	1036	RDSLCAVRTLLGYGYNIEPPDQEPQVESQSRWDRVIRFRAEKSVAVQSGRWYFEFEAV	1095		Qy	94	-----	93
Qy	36	-----TGAEILLRDP-----	44		Db	2176	ENLMTQSIGNIMNNKVFYQHPNLMRALGMHETVMEVWVNLGGESKEIRPKMVTSCCR	2235
Db	1096	TTGEMRVGNWARPRLRDPDELGADELAYVFNHGRQRWHLGSELFGRPQSGDVVGMIDL	1155		Qy	94	-----	93
Qy	45	-----	44					

```
Db 2236 FLCVFCRISRONQSMFDHLSYLLNSGIGLMOGQSTPLDVAASVIDNNELALALQEQD 2295
Qy 94 ----- 93
Db 2296 LEKVSYLAGCLOSCPMLLAKGYPDIGWPCGGERYLDLFLRFAVFVNGESVEENANVVV 2355
Qy 94 -----DP----- 95
Db 2356 RLLIRKPECFGPALRGEGGGLATIBEARISEDPAEDGQVRRDRRREHFGEEPPBEEN 2415
Qy 96 ---CHA---DLV----- 101
Db 2416 RVHLGHAIMSYAALIDLLGRCAPEMHLIQACKGEALRILRLSLVPLDVLGGIISLPL 2475
Qy 102 ----- 101
Db 2476 QIPTLGKDALVOPKMSASFPVDHKASMLVFLDRVYGIENQDFLLHLVDVGLPDMRAA 2535
Qy 102 ----- 101
Db 2536 SLDTATFTTETMALALNRYLCIAVLPLITKCAPLPAFTGTHRAIMVDSMLHTVYRLSGRS 2595
Qy 102 ----- 101
Db 2596 LTKAQRDVIECLMALCRYIRPSMLQHLRLVFDVPILNEFAKMPKLKLTNNHYERCWKY 2655
Qy 102 -----FDLEL-----PD- 113
Db 2656 YCLPTGMANFGVTSEELHLTRKLPWGIPTSLAHKKYDPELYRMAMPCLCAIAGALPDY 2715
Qy 114 -----GN----- 115
Db 2716 VDASYSKAEKATVDAEGNPDPRPVETLNVIPEKLDSEFINKEFAEYTHEKWAEDKIQNN 2775
Qy 116 ----- 115
Db 2776 WSYGENIDEELKTHPLMRPYTTFSEKDEIVRWPIKESLKAMIAEWETIEKAREGEBEKT 2835
Qy 116 ----- 115
Db 2836 EKKTRKISQSAQTYDAREGYNQPPDLSGVTLRSLOQMAEQLAENYHNTWGRKKQEL 2895
Qy 116 ----- 115
Db 2896 EAKGGTHPLLPVDTLTAKKARDKAEKQELLKELQMGVAVTRGLKDMELDTSSIEKR 2955
Qy 116 ----- 115
Db 2956 FAFGLQQLLRWMDISQEFIAHLEAVSSGRVKEKSPHEQETKFFAKILLPLINQVFTNHC 3015
Qy 116 ----- 115
Db 3016 LYFLSTPAKVLGSGGHASNKEKEMITSLFCKLAALVRHVSILFGTDAPAVNCLHLARS 3075
Qy 116 -----ROVRGVTO----- 123
Db 3076 LDARTWMSGPEIVKAGLSRFESASEIEKMVENLRGKVSQARTQVKGVGQNLTYTV 3135
Qy 124 ----- 123
Db 3136 ALLPVLTLTFOHIAHQFGDDVILDVQVSCYRTLCSYSLGTTNPNVPEKLRPALGEC 3195
Qy 124 -----LG 125
Db 3196 ARLAAMPVAFLEPQLNEYNACSVYTKSPREBAILGLPNSVEEMCPDIPVLERLMAIDG 3255
Qy 126 G----- 126
Db 3256 GLAESGARYTEMPHVIEITLPMCLSYLPRWMERGPEAPPALPAGAPPPTAVTSDHLNS 3315
Qy 127 ----- 126
Db 3316 LLGNILRIIVNNLGIDEASMMKRLAVFAQPIVSRARPELLHSHFPTIGRLRKRAKVVA 3375
```

RESULT 9

```
FAT2_RAT
ID FAT2_RAT STANDARD; PRT; 4351 AA.
AC 088277;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protocadherin Fat 2 precursor (Multiple epidermal growth factor-like domains 1).
DE FAT2 OR MEGF1.
GN Rattus norvegicus (Rat).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening.";
RL Genomics 51:27-34(1998).
CC - SIMILARITY: CONTAINS 33 CADHERIN DOMAINS.
CC - SIMILARITY: CONTAINS 1 LAMININ G-LIKE DOMAINS.
CC - SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
DR EMBL; AB011527; BAA32458.1; -
DR HSP; P08709; IBF9.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
```



```
Qy 7 ----- 6
Db 798 PPKSSWKLTVTKWMDNDNPPRPPGGYQLTISEDTEVGTITIAELKTEDADSEDNRVRV 857
Qy 7 ----- 6
Db 858 TLLTPTEKSLHPFTGELVVTGHLDRSESOYLKAEARDOPTKGHOLFVTDLIVTLIED 917
Qy 7 ----- 6
Db 918 INDNPPQITEHRRRLKVPEDMPLGTVLFLDASDDPLDPGAGEVKYLVEDAHGTFQVHPM 977
Qy 7 ----- 12
Db 978 TGALSLEKELDFERRAGYNLSFWASDSGKPLSRRTLCHVEVLVMDVNMENLHSPHSSFFV 1037
Qy 13 ----- 24
Db 1038 OQVOENSPAGTPVMVTAQDDSDGLDELQYFLRAGTGLETFSINQDTGMLTAPLADR 1097
Qy 25 ----- 24
Db 1098 EFTPYNLTVLAVDRGSVPLSAVTEVYIEVTDINDNIFSMRPPVPYPSVLEDAPLGTSVL 1157
Qy 25 ----- 24
Db 1158 QLEAWDPDSSOGKLTNLTSGNLHGFIHVHPTGTLTTAKQLDRENKDEYVLEVTVQDN 1217
Qy 25 ----- 24
Db 1218 GDSLRSTSRVVVCIILVNDNPPMPFSHKLNVRLSERLSPSPVPYRLVASDPDEGLNG 1277
Qy 25 ----- 24
Db 1278 SVTYSIESDEESFRIDPVTGVVSSSTFAAGEYNILTIKATDSCQPALSTSVRLHIEWI 1337
Qy 25 ----- 24
Db 1338 POPRSSIPLSESYSTYMETDPVNMHGVISVEGRPGLFWFHI SDGDKMDDFDIEK 1397
Qy 25 ----- 35
Db 1398 TTGSIVARPLDTRRKSSNLTVEVDGPHGTIATQVHFMIANINHRRPQFLQDHYEIRV 1457
Qy 36 ----- 42
Db 1458 PQDTLPGVLELRVQATDQDHGKGLIYTLSSODPGSANLFOLDPSSGVLVTVGTLELHSG 1517
Qy 43 ----- 42
Db 1518 PSQHILTVNRDOEMPIKRNFWVTIHVEDGNLHSPHFTQLRYEANVPDTPAGTTELLOV 1577
Qy 43 ----- 42
Db 1578 RAVDADRGAABEIHYSFLKNSDGFNIDSLIGIITVAORLYHVHLTRHALTVKAEDOGS 1637
Qy 43 ----- 42
Db 1638 PRRHDLAVIHHPSDSSAPVFSKDEYFIEIPESVPIGSPILLLSAGSSSEVTVELREG 1697
Qy 43 ----- 42
Db 1698 NKDSVFSMNSYSLISTQKRLDHEKVPSYRLRIRGSNMAGVTEVALVYIIDENDNPPA 1757
Qy 43 ----- 44
Db 1758 FGKPTFLGHISEAAPLHSLILGEDNSPLVVRASDSDREANSLLVYKILEPEALKEFKIDP 1817
Qy 45 SLGA ----- 52
Db 1818 SMGTLTITSEDFEDTLPQFNIYVHDOGTPILFAPRSKAKYIIHVRDNDSPRPFSEQIY 1877
Qy 53 ----- 52
```

```
Db 1878 EVAVVEPIHFGMGLLTVOAEDNDSRVTSYIKTSNADEAVTIHPTTGOISVVNPATLRFQ 1937
Qy 53 ----- 60
Db 1938 KFSIRASDGLYHDTAVVVKISLTVLQDKSLOFDQDVYRARTENTPHRKALVILGVGHNL 1997
Qy 61 ----- 60
Db 1998 NDTLSYFLLNGTDLFHMIESAGVLQTRGGTFDRQQDTHVAVEVRNDRVORVAQALVR 2057
Qy 61 ----- 63
Db 2058 VSEVDVNDNIPEFOHLPYTYVIOQGTPEGDVLFQVSAATDKDLGANGSVTYGAEYAYFR 2117
Qy 64 ----- 63
Db 2118 IDPYVGDISLKKPPDYQALNKYHLRIARSGIPPLQTEVEHVTVRNKSNPLFQSPYK 2177
Qy 64 ----- 63
Db 2178 VKPENITLYTPILHTQARSPEGLRLIYNIVEBEPMLFTTDFPKTVLTVTGPDYESKN 2237
Qy 64 ----- 67
Db 2238 KHVETVRATDTALGSFSEATVEVLVEDINDNPPTFSQLVYTTVSSEGSQAQTPVIQLLAS 2297
Qy 68 ----- 67
Db 2298 DODSGQNDVSYQIVEDGSDVSKFRRINGSTGEFTIQLDQYETHQHFVRKVRAMDKGDP 2357
Qy 68 ----- 67
Db 2358 PLTGETLVVNVSDINDNPPKFERPEQYEVANVSELATCGHLVLKVOALDPDIDGTSRLEYL 2417
Qy 68 ----- 80
Db 2418 ILSGNQDRHFSINSTSGIISMFNLCKQLDSSYNLRVGSQVPRATVPVYINTTNANKY 2477
Qy 81 ----- 80
Db 2478 SPEFQNVYBAELAENAKVTKVIELLAIDKDSQPGYTVDYTIINKLAGERFFINPRGOI 2537
Qy 81 ----- 80
Db 2538 TTLQKLDRENSTERVIAIKVMARDGGGKAFCTVKIILTDENDNAPQFKASGYTVSIPSN 2597
Qy 81 ----- 80
Db 2598 VSRDSPIIQVLAYDADEGRNADVTYSVDSTEDLAEIIEVNPITGVVKKVLELVLENRA 2657
Qy 81 ----- 80
Db 2658 VDFNIKAQDGGPPHWDLSLVPRLQVVPNEIPLPKFSEPLTYFSAPEDLPEGSEIGSVKAV 2717
Qy 81 ----- 80
Db 2718 AAQDPILYSLVQGTTPESNSDDVPSLDQDTGVLKVRKAMDHSTKMYQIDLMACHPHEDT 2777
Qy 81 ----- 80
Db 2778 DLVLSVSVSIQVEDVNDNRVPFEADPYKAFLTENMPGQTTVIOVTANDQDTCSDGOVSyr 2837
Qy 81 ----- 87
Db 2838 LSVPEGSNIHELFAVDSESGWITTLQELDCETQOTYRPFYVAVFDHGQTIQLSSQALVEUS 2897
Qy 88 ----- 89
Db 2898 ITDENDNPPRFASEDYRGSVVNNPGEVLATLKLADVSDQNRQVTCYITEGDPLQOF 2957
Qy 90 ----- 89
```

Db 2958 SISQGVDEWIRSRKTLDRHIAKYLLRVLTASDGKFAQSVPEVVFVVDINDNSPQCQOLL 3017
Qy 90 -----EDTDPRCH----- 97
Db 3018 YTGKVRDVTGHPILKVSADVDMDTNAQITYSLHGGAQEFKLDPHGTGELTTLTVLDR 3077
Qy 98 ----- 97
Db 3078 ERKDVYNLVAKATDGGGSCQAEVTLHIEDVNDNAPRFFPSHKCDVAVFNTVKTVPVAV 3137
Qy 98 -----ADL----- 100
Db 3138 FARDPOGANAQVVSULTDSADGQSFIDATSGVIRLEKPLQVRASSAVELTVRASDLGTP 3197
Qy 101 ----- 100
Db 3198 IPLSTLGTVTVSVGLGEDYLPFLFNAEHSHTQVPEDAPIDMEVLHLATLTPRGSEKTYGHI 3257
Qy 101 -----VLYI----- 104
Db 3258 TGGNEQCKFRLDHTGILYVNGSLDFETNPKYFLSIECSRKSSSLSDVTTIVINVTDVN 3317
Qy 105 ----- 104
Db 3318 EHHPRFTHDLYTVRVLENVAVGVILTVSASDDGPNVSAITYSLVGNQLGHFTINPKK 3377
Qy 105 ----- 104
Db 3378 GKLOVAKALDWEOTPSYSLRLRATDSGOPPLHEDTEVAEVVDVNDNPPRFFQLNYSTSV 3437
Qy 105 ----- 104
Db 3438 QENSPIGIKVLQLLDDPSPQNGPPYFFRITEGNTGSFRVTPDGLVLTAAASLKKARE 3497
Qy 105 ----- 104
Db 3498 WYQLHIEVSDGPLSLSSSTLVRVQVTEQSRYPSTPLLEISITKGEFQGMIGIHA 3557
Qy 105 ----- 104
Db 3558 TDRPDQTLTYSLEQEGGLDRYFTVGASDGKIIASOGLPHGRYSFNVTVSDGFTTTTGV 3617
Qy 105 ----- 104
Db 3618 HVHWHMEPEVPOAVNLGFHOLTPEELVSDHHRNLFRLNLLDVYKRAIHLASLOPAE 3677
Qy 105 -----TRFQLE----- 110
Db 3678 VTAGVDVLLVFERHSGTSYDLQELASAIASHVREIHSVGRIMRSALPVVPCQSQSQDQ 3737
Qy 111 ----- 110
Db 3738 TCQETVLEPRVGSYSTARLSILTTPRHILGRNCSCNGTTLRFSGQSVYQVRPLEAQNWQ 3797
Qy 111 ----- 110
Db 3798 IHFYLTQLPALLMFTNETASISLKLANGFSLHYHCPGFGYCNLSRRYPVNDQWHS 3857
Qy 111 -----LPDGNRQVRGVTL----- 124
Db 3858 LLEERDTSVHLLVDITDNASLVIPEECQGLRTERQLLLGLVPSNPSSNVSLGFEGCLDA 3917
Qy 125 -----GGACSP----- 130
Db 3918 VVNGERLELLGRKKMKRLETWALSOCCWPGTACSQSPCLNGGSCSPALGSGYLCRCP 3977
Qy 131 ----- 130
Db 3978 PPSGRNCELGRNCTSAQCQEGGTCVSSPEGTSCNCPHYTGDCRMEARGCGGHCLI 4037
Qy 131 ----- 130
Db 4038 TPEIKRGDWGQOEFLVITVALPLVIATVGLLLYCRRRKSHKPVMTMEDPDLARSIGVDT 4097

Qy 131 ----- 130
Db 4098 QASPAIELOPLNTSSNNLNQPEPSKTSVPNELVTGPPSSKORPMVCSVPPRLPPAAVSS 4157
Qy 131 -----TWS-----CL 135
Db 4158 HPGHEPIIKRTWSGEELVYPSGAAVMPPTYSRKKHWEYHPHETMOGTLPPSPRRHVGPAV 4217
Qy 136 ITEDTGF-----DL----- 144
Db 4218 MPDPTGLYGGFPFPLELENKRAPLPPTYGNQNLDELMPRPSPREHLLAPCLNNEYTAIS 4277
Qy 145 -----GVTI 148
Db 4278 YHSPQROGGGGPCLAEAGYKGVSM 4302
RESULT 10
RYR1 HUMAN
ID RYR1 HUMAN STANDARD; PRT: 5038 AA.
AC P21817; Q16314; Q9NPK1; Q9PLU4;
DT 01-MAY-1991 (Rel. 18, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ryanodine receptor 1 (Skeletal muscle-type ryanodine receptor) (RYR1)
DE (RYR-1) (Skeletal muscle calcium release channel).
GN RYR1 OR RYDR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND PARTIAL SEQUENCE.
RX TISSUE=Skeletal muscle;
RX MEDLINE=90130482; PubMed=2298749;
RA Zorzato F., Fujii J., Otsu K., Phillips M.S., Green N.M., Lai F.A.,
RA Meisner G., MacLennan D.H.;
RT "Molecular cloning of cDNA encoding human and rabbit forms of the
RT Ca2+ release channel (ryanodine receptor) of skeletal muscle
RT sarcoplasmic reticulum";
RL J. Biol. Chem. 265:2244-2256(1990).
RN [2]
RP REVISIONS TO 2324; 2840 AND 3380, VARIANT MH ARG-248, AND VARIANTS
RP CYS-471; LEU-1787 AND CYS-2060.
RC TISSUE=Muscle;
RX Gillard E.F., Otsu K., Fujii J., Duff C.L., de Leon S., Khanna V.K.,
RA Britt B.A., Morton R.G., McLennan D.H.;
RT "Polymorphisms and deduced amino acid substitutions in the coding
RT sequence of the ryanodine receptor (RYR1) gene in individuals with
RT malignant hyperthermia";
RL Genomics 13:1247-1254(1992).
RN [3]
RP REVISIONS TO 1365-1368, VARIANT CCD H-2435, AND ALTERNATIVE SPLICING.
RC TISSUE=Muscle;
RX MEDLINE=94035117; PubMed=8220422;
RA Zhang Y., Chen H.S., Khanna V.K., de Leon S., Phillips M.S.,
RA Schappert K., Britt B.A., Brownell A.K.W., McLennan D.H.;
RT "A mutation in the human ryanodine receptor gene associated with
RT central core disease";
RL Nat. Genet. 5:46-50(1993).
RN [4]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=96299657; PubMed=8661021;
RA Phillips M.S., Fujii J., Khanna V.K., de Leon S., Yokobata K.,
RA de Jong P.J., McLennan D.H.;
RT "The structural organization of the human skeletal muscle ryanodine
RT receptor (RYR1) gene";
RL Genomics 34:24-41(1996).
RN [5]
RP SEQUENCE OF 598-722 FROM N.A.
RC TISSUE=Skeletal muscle;

RX MEDLINE=92347887; PubMed=1639409;
RA Otsu K., Phillips M.S., Khanna V.K., de Leon S., McLennan D.H.;
RT "Refinement of diagnostic assays for a probable causal mutation for
RT porcine and human malignant hyperthermia.";
RL Genomics 13:835-837(1992).
RN [16]
RP SEQUENCE OF 3339-5038 FROM N.A. (ISOFORM 2).
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala N., Terry A., Ganes J.,
RA Dangnan L., Eler A., Christensen M., Georgescu A., Avila J.,
RA Liu S., Attix C., Andreise T., Trankheim M., Amico-Keller G.,
RA Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
RA Krommiller B., Agellano A., Saunders C., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [17]
RP SEQUENCE OF 4696-4974 FROM N.A.
RC TISSUE=Myometrium;
RX MEDLINE=96032536; PubMed=7556644;
RA Lynn S., Morgan J.M., Lamb H.K., Meissner G., Gillespie J.I.;
RT "Isolation and partial cloning of ryanodine-sensitive Ca²⁺ release
RT channel protein isoforms from human myometrial smooth muscle.";
RL FEBS Lett. 372:6-12(1995).
RN [18]
RP TISSUE SPECIFICITY.
RX MEDLINE=98268728; PubMed=9607712;
RA Martin C., Chapman K.E., Seckl J.R., Ashley R.H.;
RT "Partial cloning and differential expression of ryanodine
RT receptor/calcium-release channel genes in human tissues including
RT the hippocampus and cerebellum.";
RL Neuroscience 85:205-216(1998).
RN [19]
RP S-NITROSYLATION OF CYS-3635.
RX MEDLINE=21457268; PubMed=11562475;
RA Sun J., Xin C., Eu J.P., Stampler J.S., Meissner G.;
RT "Cysteine-3635 is responsible for skeletal muscle ryanodine receptor
RT modulation by NO.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:11158-11162(2001).
RN [20]
RP VARIANT MH CYS-614.
RX MEDLINE=92128959; PubMed=1774074;
RA Gillard E.F., Otsu K., Fujii J., Khanna V.K., de Leon S.,
RA Derdemizi J., Britt B.A., Duff C.L., Worton R.G., McLennan D.H.;
RT "A substitution of cysteine for arginine 614 in the ryanodine
RT receptor is potentially causative of human malignant hyperthermia.";
RL Genomics 11:751-755(1991).
RN [21]
RP VARIANTS CCD CYS-163 AND MET-403.
RX MEDLINE=94035118; PubMed=8220423;
RA Quane K.A., Healy J.M.S., Keating K.E., Manning B.M., Couch F.J.,
RA Palmucci L.M., Doriguzzi C., Fagerlund T.H., Berg K., Ordling H.,
RA Bendixen D., Mortier W., Linz U., Muller C.R., McCarthy T.V.;
RT "Mutations in the ryanodine receptor gene in central core disease and
RT malignant hyperthermia.";
RL Nat. Genet. 5:51-55(1993).
RN [12]
RP VARIANT MH SER-522.
RX MEDLINE=95130087; PubMed=7829078;
RA Quane K.A., Keating K.E., Healy J.M.S., Manning B.M.,
RA Krivosic-Horber R., Krivosic I., Monnier N., Lunardi J.,
RA McCarthy T.V.;
RT "Mutation screening of the RYR1 gene in malignant hyperthermia:
RT detection of a novel Tyr to Ser mutation in a pedigree with
RT associated central cores.";
RL Genomics 23:236-239(1994).
RN [13]
RP VARIANT MH ARG-341.
RX MEDLINE=94282042; PubMed=8012359;
RA Quane K.A., Keating K.E., Manning B.M., Healy J.M.S., Monsieus K.,
RA Heffron J.J.A., Lehane M., Heytens L., Krivosic-Horber R., Adnet P.,
RA Ellis F.R., Monnier N., Lunardi J., McCarthy T.V.;
RT "Detection of a novel common mutation in the ryanodine receptor gene

RT in malignant hyperthermia: implications for diagnosis and
RT heterogeneity studies.";
RL Hum. Mol. Genet. 3:471-476(1994).
RN [14]
RP VARIANT MH ARG-2434.
RX MEDLINE=95152512; PubMed=7849712;
RA Keating K.E., Quane K.A., Manning B.M., Lehane M., Hartung E.,
RA Censier K., Urwyler A., Klausnitzer M., Muller C.R., Heffron J.J.A.,
RA McCarthy T.V.;
RT "Detection of a novel RYR1 mutation in four malignant hyperthermia
RT pedigrees.";
RL Hum. Mol. Genet. 3:1855-1858(1994).
RN [15]
RP VARIANT MH ARG-2434.
RX MEDLINE=95187158; PubMed=7881417;
RA Phillips M.S., Khanna V.K., de Leon S., Frodis W., Britt B.A.,
RA McLennan D.H.;
RT "The substitution of Arg for Gly2433 in the human skeletal muscle
RT ryanodine receptor is associated with malignant hyperthermia.";
RL Hum. Mol. Genet. 3:2181-2186(1994).
RN [16]
RP VARIANT MH CYS-614.
RC TISSUE=Blood;
RX MEDLINE=95271229; PubMed=7751854;
RA Moroni I., Gonano E.F., Comi G.P., Tegazzin V., Prella A., Bordini A.,
RA Bresolin N., Scarlato G.;
RT "Ryanodine receptor gene point mutation and malignant hyperthermia
RT susceptibility.";
RL J. Neurol. 242:127-133(1995).
RN [17]
RP VARIANT MH ARG-35.
RX MEDLINE=97219028; PubMed=9066328;
RA Lynch P.J., Krivosic-Horber R., Reyford H., Monnier N., Quane K.A.,
RA Adnet P., Haudecoeur G., Krivosic I., McCarthy T.V., Lunardi J.;
RT "Identification of heterozygous and homozygous individuals with the
RT novel RYR1 mutation Cys35Arg in a large kindred.";
RL Anesthesiology 86:620-626(1997).
RN [18]
RP VARIANT MH LEU-614.
RX MEDLINE=98051290; PubMed=9389851;
RA Quane K.A., Ordling H., Keating K.E., Manning B.M., Heine R.,
RA Bendixen D., Berg K., Krivosic-Horber R., Lehmann-Horn F.,
RA Fagerlund T.H., McCarthy T.V.;
RT "Detection of a novel mutation at amino acid position 614 in the
RT ryanodine receptor in malignant hyperthermia.";
RL Br. J. Anaesth. 79:332-337(1997).
RN [19]
RP VARIANT MH TRP-552.
RX MEDLINE=97284075; PubMed=9138151;
RA Keating K.E., Giblin L., Lynch P.J., Quane K.A., Lehane M.,
RA Heffron J.J.A., McCarthy T.V.;
RT "Detection of a novel mutation in the ryanodine receptor gene in an
RT Irish malignant hyperthermia pedigree: correlation of the IVCT
RT response with the affected and unaffected haplotypes.";
RL J. Med. Genet. 34:291-296(1997).
RN [20]
RP VARIANTS MH CYS-2163; HIS-2163; MET-2168 AND MET-2206.
RX MEDLINE=98163444; PubMed=9497245;
RA Manning B.M., Quane K.A., Ordling H., Urwyler A., Tegazzin V.,
RA Lehane M., O'Halloran J., Hartung E., Giblin L.M., Lynch P.J.,
RA Vaughan P., Censier K., Bendixen D., Comi G.P., Heytens L.,
RA Monsieus K., Fagerlund T.H., Wolz W., Heffron J.J.A., Mueller C.R.,
RA McCarthy T.V.;
RT "Identification of novel mutations in the ryanodine-receptor gene
RT (RYR1) in malignant hyperthermia: genotype-phenotype correlation.";
RL Am. J. Hum. Genet. 62:595-609(1998).
RN [21]
RP VARIANTS MH CYS-2459 AND HIS-2459.
RX MEDLINE=98111378; PubMed=9450902;
RA Manning B.M., Quane K.A., Lynch P.J., Urwyler A., Tegazzin V.,
RA Krivosic-Horber R., Censier K., Comi G.P., Adnet P., Wolz W.,
RA Lunardi J., Muller C.R., McCarthy T.V.;
RT "Novel mutations at a CpG dinucleotide in the ryanodine receptor in

```
.RT malignant hyperthermia. ";
Query Match 27.2%; Score 210; DB 1; Length 5038;
Best Local Similarity 2.5%; Pred. No. 64;
Matches 81; Conservative 33; Mismatches 34; Indels 3101; Gaps 26;

Qy 1 AAGGILHL----- 8
Db 555 ASSGILEVLCVLTIESPEVLNIIQENHIKSIISLLDKHGRNHKVLVDVLCVLCVNGVAVR 614
Qy 9 -----ELLV-----AVGPDVFOAHQEDTERY----- 29
Db 615 SNODLITENLPGRELLQTNLINVTSIRDNI FVGRAEGTQYSKWYFVMVDEVTPFL 674
Qy 30 --VLTNINIG----- 37
Db 675 TAOATHLRVGVWALTGYTPYPGAGEGMGNGVGGDDLYSGFDGLHWTGHVAREVPTSPGQ 734
Qy 38 ----- 37
Db 735 HLLAPEDVISCCDLSPVSTISFRINGCPVQGVFESFNLDGLFPFVVSAGVKVRFLLGG 794
Qy 38 ----- 37
Db 795 RHGEFKFLPPGYAPCHEAVLPRERLHLEPIKEYRREGPRGPHLVGPSRCLSHTFVPCP 854
Qy 38 ----- 37
Db 855 VDTQIVLPHPLERIREKLAENIHELMWALTRIEQGWTVGPVRDDNKRLLHPCLVDFHSLPE 914
Qy 38 ----- 37
Db 915 PERNYNLQMSGETLKTLLALCHVGMADKAEDNLKTKLPKTYMNSNGYKAPAPLDLSHV 974
Qy 38 ----- 37
Db 975 RLTPAQTTLVRLAENGHNWAVDRVGQGSYSYSAVQDIPARRNPRLVPYRLDDEATKRSN 1034
Qy 38 ----- 37
Db 1035 RDSLQAVRTLGGYNIETPPDQPSQVENQSRCDVRIFRAEKSYTVQSGRWYFEFAV 1094
Qy 38 -----AELLRDSLGA----- 48
Db 1095 TTGEMRVGWARPELRPDVELGADELAYVFNCHRGORHLGSEPPGRPWQPDVVGCMIDL 1154
Qy 49 ----- 48
Db 1155 TENTIIFTLNGEVLMSDSGSETAPREIEIGDGLPVCGLPGQVGHNLNGQDVSSLRFFA 1214
Qy 49 ----- 48
Db 1215 ICGLOEGPEPFAINMQRPVTTWFSKGLPQFEPVPLEHPHYEVSVDGTVDTPPCLRLTHR 1274
Qy 49 -----OFRVHL----- 54
Db 1275 TWGSONSLVEMFLRLSLPVOFHQFRCTAGATPLAPGLOPPADEARAAEPDPDYENL 1334
Qy 55 -----VK 56
Db 1335 RRSAGGWEAENGKEGTAKGAPGCTPOAGGEAQPARAENKDATTEKNKKRGFLFKAKK 1394
Qy 57 MVILTEPEGAPNI----- 69
Db 1395 VAMMTQPPATPTLPRPHDVVPADNRDDPEIILNTTYYYSRVRFAGQEPSCVMAGWVTP 1454
Qy 70 ----- 69
Db 1455 DYHQHMSFDSLKVRVTVTWGDEQGNVHSSLKSCNCTWVGDPVSPGQQRISHTDLV 1514
Qy 70 -----TAN----- 72
Db 1515 IGLVDLATGLMTFTTANGKESNTFFQVBENTKLPFAVFLPTHQNVIQFELGKQKNIMPL 1574
```

QY 102 -----LVYTR-----FDLEL-----PD- 113
DB 2655 YCLPTGWANFGVTSBEELHLTRKLFWGFDSLAHKYDPELYRMAMPCLCAIAGALPPDY 2714
QY 114 -----GN----- 115
DB 2715 VDASYSSKAEKKATVDAEGNFDPRPVETLNVIIPEKLDSEFNKFAEYTHEKWAFDKIQNN 2774
QY 116 ----- 115
DB 2775 WSYGENIDEELKTHPMLRPYKTFSEKDKEIYRPIKESLKAMIANEWETIEKAREGESEKT 2834
QY 116 ----- 115
DB 2835 EKKKTRKISQSAQTYDPREGYNPQPDLSAVTSLRELOQAELAENVHNTWGRKKQBEL 2894
QY 116 ----- 115
DB 2895 EAKGGGTHPLLVYDPTLTAKEKARDREKAQELLKFLQNMNGYAVTRGLKDMELDSSIEKR 2954
QY 116 ----- 115
DB 2955 FAFGLQQLLRHMDISOEFIAHLEAVVSSGRVEKSPHEQEIKFPAKILLPLINQYFNHC 3014
QY 116 ----- 115
DB 3015 LVFLSTPAKVLGGSHASNKEKEMITSLFCKLAALVRHRVSLFGTDAPAVVNCILHARS 3074
QY 116 -----RQVRGVTO----- 123
DB 3075 LDARTVMKSGPEIVKAGLRSPFESADIEKVENLRGLKGVSQARTQVKGQONLTYTTV 3134
QY 124 ----- 123
DB 3135 ALLPVLVTLFOHIAHQFGDDVLDQVQSVYRTLCISYSLGTTKNTYVEKURPALGECL 3194
QY 124 -----LG 125
DB 3195 ARLAAAMPVAFLEPQLEYNACSVYTKSPRERAILGLPNSVEEMCPDIPVLERLWADIG 3254
QY 126 G----- 126
DB 3255 GLAEGARYTEMPHVIEITPLMCLSYLPRWERGPEAPPALPAGAPPCTAVTSDHLNS 3314
QY 127 ----- 126
DB 3315 LLGNLRIIVNGLIDEASMMKRLAVFAQPIVSRARPELLOSHFPTIGRLRKRAGKVV 3374
QY 127 ----- 126
DB 3375 EEEQLRLEAKAEOGELLVRDEFVLCRDLYALYPLLIRYVYDNNRAQWLTEPNPSABEL 3434
QY 127 ----- 126
DB 3435 FRMVGEIFIYWSKSHNFKEQNFVQNEINNMSTADNKSMAKAGDIOSGGSDQERT 3494
QY 127 -----ACSP----- 131
DB 3495 KKKRGDRYSVQTSIVATLKKMLPIGLNMCAPTQDQLITLAKTRYALKOTDEEVREFLH 3554
QY 132 ----- 131
DB 3555 NNHLQGVKESPLRWQVALYRGVPGREEDADDPEKIVRRVQEVSAVLYLDQTEHPYK 3614
QY 132 -----W-----SC-----LITEDTGF----- 143
DB 3615 SKKAVVHKLLSKORRRAVACFRMTPLYNLPHTRACNMFLSYKAAWILTEDHSFEDRW 3674
QY 144 ----- 143
DB 3675 DDLSKAGEQEEEEVEEKPPDLHLQLVLFHSRTALTEKSLDEDLYMAYADIMAKSCH 3734
QY 144 ----- 143

DB 3735 LEEGGENGEABEEVEVSFEKQMEKQRLLYQOARLHTRGAABMVLQMIASACKGETCAMVS 3794
QY 144 ---LGVTI 148
DB 3795 STLKLGIS 3803

RESULT 11

TYCC_BACBR
ID TYCC_BACBR STANDARD; PRT; 6486 AA.
AC O30409;
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Tyrocidine synthetase III [Includes: ATP-dependent asparagine
adenylase (AsnA) (Asparagine activase); ATP-dependent glutamine
adenylase (GlnA) (Glutamine activase); ATP-dependent tyrosine
adenylase (TyrA) (Tyrosine activase); ATP-dependent valine adenylase
(ValA) (Valine activase); ATP-dependent ornithine adenylase (OrnA)
(Ornithine activase); ATP-dependent leucine adenylase (LeuA) (Leucine
activase)].
GN TYCC
OS Bacillus brevis.
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX NCBI_TaxID=1393;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 8185;
RX MEDLINE=98012987; PubMed=9352938;
RA Mootz H.D., Marahiel M.A.;
RT "The tyrocidine biosynthesis operon of Bacillus brevis: Complete
nucleotide sequence and biochemical characterization of functional
internal adenylation domains".
RL J. Bacteriol. 179:6843-6850(1997).
CC -!- FUNCTION: INCORPORATES SIX AMINO ACIDS (FOR TYROCIDINE A, ASN,
GLN, TYR, VAL, ORN, AND LEU) IN THEIR L-CONFIGURATION INTO THE
PEPTIDE PRODUCT.
CC -!- COPACTOR: CONTAINS 6 COVALENTLY BOUND PHOSPHOPANTHETHEINES (BY
SIMILARITY)
CC -!- PATHWAY: Cyclic peptide antibiotic tyrocidine biosynthesis.
CC -!- SUBUNIT: LARGE MULTIENZYME COMPLEX OF TYCA, TYCB AND TYCC.
CC -!- DOMAIN: CONSISTS OF SIX MODULES, AND HARBORS A PUTATIVE
THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE
INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE
FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE
ADENYLATION, THIOYLATION, CONDENSATION (NOT FOR THE INITIATION
MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION
(OPTIONAL).
CC -!- MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC
DECAPEPTIDES, TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL-
ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR
RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE
CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.
CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
FAMILY.
CC -!- SIMILARITY: CONTAINS 6 ACYL CARRIER DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.ebi.ac.uk/ebis-sib.ch/>)
or send an email to license@sib-sb.ch.
CC
CC EMBL; AF004835; AAC45930.1; -.
DR HSSP; P14687; 1AMU
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001242; Condensatn.
DR InterPro; IPR003880; Pantine attach.
DR InterPro; IPR000379; Ser estase site.
DR InterPro; IPR001031; Thioesterase.

```

DR Pfam; PF00501; AMP-binding; 6.
DR Pfam; PF00550; DP-binding; 6.
DR Pfam; PF00668; Condensation; 6.
DR Pfam; PF00975; Thioesterase; 1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 6.
DR PROSITE; PS00455; AMP BINDING; 6.
DR PROSITE; PS00075; ACP_DOMAIN; 6.
DR KW Ligase; Antibiotic biosynthesis; Phosphopantetheine;
KW Multifunctional enzyme; Repeat.
FT REPEAT 466 1038 DOMAIN 1 (ASPARAGINE-ACTIVATING).
FT REPEAT 1521 2070 DOMAIN 2 (GLUTAMINE-ACTIVATING).
FT REPEAT 2536 3113 DOMAIN 3 (TYROSINE-ACTIVATING).
FT REPEAT 3590 4149 DOMAIN 4 (VALINE-ACTIVATING).
FT REPEAT 4606 5203 DOMAIN 5 (ORNITHINE-ACTIVATING).
FT REPEAT 5658 6245 DOMAIN 6 (LEUCINE-ACTIVATING).
FT DOMAIN 970 1037 ACYL CARRIER (ACP) 1.
FT DOMAIN 2007 2074 ACYL CARRIER (ACP) 2.
FT DOMAIN 3045 3112 ACYL CARRIER (ACP) 3.
FT DOMAIN 4080 4147 ACYL CARRIER (ACP) 4.
FT DOMAIN 5124 5191 ACYL CARRIER (ACP) 5.
FT DOMAIN 6167 6234 ACYL CARRIER (ACP) 6.
FT BINDING 1000 1000 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 2037 2037 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 3075 3075 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 4110 4110 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 5154 5154 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 6197 6197 PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ SEQUENCE 6486 AA; 724011 MW; 4934900AF0DF786 CRC64;

Query Match 27 28; Score 210; DB 1; Length 6486;
Best Local Similarity 1.3%; Pred. No. 1.3e+02;
Matches 78; Conservative 32; Mismatches 33; Indels 6044; Gaps 23;

QY 3 GGILH----- 7
DB 152 GVLLHLKLLTYAALRKGEPIPREATKPYSEYIKWLDKQNKDEALAYQWYLAGYDHOAAF 211
QY 8 ----- 7
DB 212 PKKLGTEASRYEHVEAMFTIAPEKTQOLIQIANQONATSSVFQALWGILASTYKNADD 271
QY 8 ----- 7
DB 272 VVFGSVSGRRPQIQIESVMVGLFINTIPTRVQTNKQOTFSELLQTVQKQALASATYDFA 331
QY 8 ----- 7
DB 332 PLYEIQTTVLKOELIDHLVTFENYPDHSMKHEESLGFQFTVESGDEQTSYDLNVVVAL 391
QY 8 ----- 7
DB 392 APSNELYVKSNAAYSESVNRIEGHLRTVIDQVIGNPHVHLHEIGIITEEBEQQLLV 451
QY 8 ----- 7
DB 452 AYNTAAEYPRDKTIFELIAEQASRTPAKAAVCGEDTLTYQELMERSAQLANALREKGI 511
QY 8 -----LELVAV----- 14
DB 512 ASGSIVSMAEHSLELIVAMVLRSGAAYLPIDPEYQDRIQYLLDDSDQTLTLLTOSHL 571
QY 15 ----- 14
DB 572 QPNIRFAGSVLYLDDRLSYEGGSTFAPESKPDDLAYMIYVTSGTGNPKGAMITHQGLVN 631
QY 15 ----- 14
DB 632 YIWANKVYQGEAVDPFLYSISFDLTVTSTIFTPLLSGNTIHVYRGADKVQVILDIKD 691
QY 15 ----- 14
DB 692 NKVGIIKLTPHLKLIHIDGKASSIRRFIVGGENLPTKLAKQIYDHFGENVQIFNEYGP 751

```

```

QY 15 ----- 14
DB 752 TETVGCMIYLYDPQTTOESVPFGVPADNVQLYLLDASMPQVPVPSGLGEMYIAGDGVAK 811
QY 15 ----- 14
DB 812 GYFNRPBELTKKEFIDNPFPRGTYMYRTGDLAKWLPDGNMEYAGRMDOYKIRGRHIEGE 871
QY 15 ----- 14
DB 872 IETRLTQHEAVKEAVIVEKDESGQNVLYAYLVSERELTVAELEREFLGRTLPSYMIPSFF 931
QY 15 ----- 14
DB 932 IRLAEIPTANGVERKKLPKAGAVVTGTAYAAPQNEIEAKLAEIMQVVLGISQVGIHD 991
QY 15 ----- 14
DB 992 DFFDLGGHSUKAMTVVQVSKALEVELPVKALFEHPTVAELARFLSRSEKTEYTAIOPVA 1051
QY 15 ----- 14
DB 1052 AQEFYPVSSAQRMYILQQFEGNGISYNSGAILLEGKLDYARFASAVQQLAERHEALRT 1111
QY 15 ----- 14
DB 1112 SFHRIDGEPVQVHEEVEVPLFMLEAPEQAEKIMREFVRPFDGLGVAPLMRTGLLLGKGD 1171
QY 15 ----- 14
DB 1172 RHLFLDMHHIISGVSSQILLREFAELYOGADQLPSLQYKDYKQFAAQNELFOTEAYKKQ 1231
QY 15 ----- 14
DB 1232 EQHWNLTFADEIPLNLPTDYPSPVSQSPAGDLVLPFAAGKELLERLQOVASETGTLYMI 1291
QY 15 ----- 14
DB 1292 LLAAYNVLLSKYTGQEDIIIVGTPVAGRSHADVENIMGIFVNTLALRNQPASSKTFQAFLO 1351
QY 15 ----- 14
DB 1352 EVKQNALAAAYDHODYPFEEELVEKLAIORISRNPLFDTLFSLENANQOOLAIABLTASPY 1411
QY 15 ----- 14
DB 1412 ELFNKISKFDLALNASESPADIQFLTATKLPFKKETVERMARHYLEILRWISEOPTASL 1471
QY 15 ----- 14
DB 1472 ADIDMMTEAKRTLNLNVNDTFVERTAATLHQLVBEQQAARTPDEVAVVVEVALTYREL 1531.
QY 15 -----GPDV----- 18
DB 1532 NAPANOLARLLRSHGTGPDTLIGIMYDRSPGMVGLAVLKAGGAYTPIDPSYPERIOY 1591
QY 19 ----- 18
DB 1592 MLSDSOAPILLTQRHLQELAAAYQGEIIDVEEAITYGADTNLDNVAGKODLAYVIYVTSGS 1651
QY 19 ----- 18
DB 1652 TGNPKGWMISHQAI CNHMLMWMRETPLTTTDAVLQKTPFSFDASVMEFVPLITGGOLV 1711
QY 19 ----- 18
DB 1712 AKPDGHRDIA YMTRLIRDEKITTLQWMPVSLDLVMTDPGWSACTSLQRFVCGGEALTAL 1771
QY 19 ----- 18
DB 1772 VSRFYETQQAQLINLYGPTTETTDATYWPFCPRQOEYSATIPGKPIDNVRLYVNVASNQLO 1831

```

Qy	19	-----	18
Db	1832	PVGAGELCIAGDGLARGYQREBELTKASFDVDPFPGGTYRTGDMVRLPDGHIEYLG	1891
Qy	19	-----	18
Db	1892	RIDHOVKIRHRIELGEIEATLQHEAVKAVVMARQDGQNSLYAVVABODIQTAEL	1951
Qy	19	-----	18
Db	1952	RTYLSATLPAYMVP\$AFVLEQLPLSANGKVRKALPQPEDAAASAAVVPAPRNEWAKL	2011
Qy	19	-----	18
Db	2012	AAIWESVLGVEPIGVGHDHFFELGHSLKAMHVISLLQRSFQVDVPLKVLFSPTIAGLAP	2071
Qy	19	-----	18
Db	2072	LVAARKGTYTAIPPEKEQYYPV\$AAQKRMFILOQMEGAGISYNNPGFMYLDGKLDTER	2131
Qy	19	-----	18
Db	2132	LQQAOKSLVORHESLRTSPHSVOGETVQRVHDDVDLAI\$FGTEATEAQIAEQFIQPF	2191
Qy	19	-----	18
Db	2192	LCTAPLLRAGLIKLAPERHFLMDLHHI\$VVDGVSIGLLIEBFAQLYHGEELPALRIQYK	2251
Qy	19	-----	18
Db	2252	FAKQODWFOTEBFAQOAYWLNTFTGEI\$PVLNLPDYP\$P\$VK\$FAGDREVFVSGTALP	2311
Qy	19	-----	18
Db	2312	KQLHQAQETGTYLWVLLAAYNVLLSKYSROEDII\$VGAPTAGRSHAETESIVGMFNTL	2371
Qy	19	-----	18
Db	2372	ALRNEPAGGKTRFDFLAEVKINTLGA\$FHQDYP\$LDLVKDMQDLSRNP\$P\$TVFILQ	2431
Qy	19	-----	27
Db	2432	NMEOKPFEMEQLTITP\$AEVKQAKFDLSLEAYEENAEI\$IFSLDY\$TKLFSRETIEKIAT	2491
Qy	28	-----	27
Db	2492	HFIQILRAVIAEBEMPLSEITMLTEAKORLLVD\$P\$NGAHKOP\$P\$NKTLOALFEEQAEKSP	2551
Qy	28	-----	27
Db	2552	QATAVEISGQPLSYQELNERANQLAATLRERG\$VQDQ\$P\$VGIMANRSVEMVVGILAILKAG	2611
Qy	28	-----	40
Db	2612	GAYVPIDPEYPERVAYMLTDCOARLVLTQ\$H\$KAGLGS\$V\$TAECLYLDDESNGVGHRSN	2671
Qy	41	-----	40
Db	2672	LQ\$INTASDLAYIIT\$G\$TGKPGVMVEHRGI\$VNNVLW\$KAEYQ\$M\$KVGDR\$LL\$LS\$FAP	2731
Qy	41	-----	44
Db	2732	DAFVLSF\$P\$TVL\$GATV\$VLAED\$EAKD\$P\$V\$KLK\$IAASR\$CT\$MTG\$V\$P\$SL\$FQ\$AL\$EC\$TPA	2791
Qy	45	-----	49
Db	2792	DIRPLOT\$VTL\$GGEK\$IT\$AQL\$VEK\$CQ\$LN\$PDL\$VIV\$N\$EY\$G\$P\$T\$ESS\$V\$VAT\$Q\$R\$LAG\$P\$DA\$AIT\$IG	2851
Qy	50	-----	49
Db	2852	RPIANTS\$LYI\$N\$Y\$H\$Q\$LO\$PI\$GV\$VEI\$CIGGR\$GLARG\$Y\$N\$K\$PAL\$TEE\$K\$F\$V\$SH\$P\$FA\$G\$R\$MY	2911
Qy	50	-----	49

Db	2912	KTGDLGKWLPGDTIEYIGRIDEQVKRGYRIEIGEIESALLAAABKUTAAVVVYVEDOLQG	2971
Qy	50	-----	49
Db	2972	SALAAFTADEQLDVTKLMSHLSKRLPSYMI PAHFVQLDOLPLTPNGKVDKKALPKPEGK	3031
Qy	50	-----	49
Db	3032	PVTEAQYVPTNAVESKLAEIWERVLGSGILDNFFQJGGHSLKAMAVAAQVHREYQV	3091
Qy	50	-----	49
Db	3092	ELPLKVLFAOPTIKALAAQVATSGKETYPVIEPAPLQEYYPVSSAQKRMVYLRFADTGT	3151
Qy	50	-----	49
Db	3152	VYNMPSALYIEGDLDRKRFEAALHGLVERHESLRTSFHTVNGEPVORVHEHVELNVOYAE	3211
Qy	50	-----	49
Db	3212	VTEAQVEPTVESFVQAFDLTKAPLLRVGLFKLAARKHFLDMDHHIISDGVSAGIIMEEF	3271
Qy	50	-----	49
Db	3272	SKLYRGEELPALSVHYKDPFVQSELFSQSDVYTEHENYWLNAFSGDIPVLNLPADFSRPL	3331
Qy	50	-----	49
Db	3332	TQSPGDCVSPQADKALLDDLHKLQAESOSTLPMVLLAAYNVLLAKYSGOEDIVWGTPIA	3391
Qy	50	-----	49
Db	3392	GRSHADIENVLGMFNTLALRNPVETKHFQAFLEBVKQNTLOAYAHODYPFEALVEKLD	3451
Qy	50	-----	49
Db	3452	IORDLSRNPFLTMTFILQNLQKAYELDGLKLEAYPAQAGNAKFOLDLEBAHEDETGIFHA	3511
Qy	50	-----	49
Db	3512	LVYSTKLFORESIERMAGHFLQVLROQVADQATALREISLSEBERRIVTVDFNNTFAAY	3571
Qy	50	-----	49
Db	3572	PRDLTIQELFEOQAAKTPEHAAVMDGQMLTYRELNEKANQALHVLQNGVGKESIVGLL	3631
Qy	50	-----	49
Db	3632	ADRSLEMITGILKAGAYLGLDPEHPERLAYMLEDGGVKVVLVQKHLPLVGEGLM	3691
Qy	50	-----	49
Db	3692	PIVLEESLRPEDCGNPAIVNGASDLAYMYMTSGTGPKGMVEHRNVTRLVMHTNYVQ	3751
Qy	50	-----	49
Db	3752	VRESDRMIQTGAIGFDAMTFEIFGALLHGASLYLVSKDVLDAEKLGDFLRTNOITTMWL	3811
Qy	50	-----	49
Db	3812	TSPLFNQLSQDNPAFDSLRALI VGEGLSPKHINRVKSALPDLEIWNGYGPENTTFST	3871
Qy	50	-----	49
Db	3872	CYLIEQHFEEQITPIGKPIANSTAYIVDGNNOPOPIGVPGLCVGGDGVARGVYNKPDELTA	3931
Qy	50	-----	49
Db	3932	EKFVPNPFPAGETMYRTGDARWLPDGTIEYLGRIDQQVKIRGYRIELGEIETVLSQOAO	3991
Qy	50	-----	49

```
Db 3992 VKEAVVAIEEANGOKALCAFYVPEQAVDAEALREAMSKQLPGYMWVPAYYVQMEKLPITA 4051
Qy 50 ----- 49
Db 4052 NGKVDRRALPOPSGERTTGSFAVAQNDTEAKLOQIWOEVILGIPAIGHDNFFBIGHSL 4111
Qy 50 ----- 49
Db 4112 KAMNVITQVHKTFQVELPLKALFATPTIHELAAHIAESAFQFETIQPVEPAAPYPSFA 4171
Qy 50 ----- 49
Db 4172 QKRWYILHQFEGSGISYNVPSVLVEGKLDYDRFAAAIQSLVKRHSRTSFHSGVNGEPL 4231
Qy 50 ----- 56
Db 4232 QRVPDVELPVRLLTEATEDQESLIQELIOPDLEIAPLFRVNLIKLGAERHLFFMDMH 4291
Qy 57 ----- 56
Db 4292 IISGVSLAVIEEIASLYAGKQLSDLRIQYKDFAVWQTKLAQSDRFQKQEDFWTRTFAG 4351
Qy 57 ----- 61
Db 4352 EIPLNLPHDYPRPSVQSGDGTVALGCTGHHLEQLKLAETGTTFLFWLLAAHYVLLS 4411
Qy 62 ----- 61
Db 4412 KYAQEEIVGTPIAGRSHADVERIVGMFVNTLAKNTAAGSLSFRAFLEDVKQNALHAF 4471
Qy 62 ----- 61
Db 4472 EHODYPEHLVEKLOVRDLRSNPLFDTMFSGLAESAEVADLKVSPVPVNGHIAKFD 4531
Qy 62 ----- 61
Db 4532 LSLDAMEKODGLLVQFSYCTKLFAKETVDRLLAAHYVQLLOITADPDIELARISVLSKAE 4591
Qy 62 ----- 61
Db 4592 TEHMLSHFLATKATYPTDKTFQKLFEEQVEKTPNEIAVLFNGEQLTYOELNAKANQLARV 4651
Qy 62 ----- 61
Db 4652 LRRKGVPKSTVGILVDRSLYVMGLAVLKAGGTVPIDPDYPLERQAPMLEDSEAKLL 4711
Qy 62 ----- 66
Db 4712 LTLOKMSOAVFPYETFYLDTETVDQEBETGNLHVAPQENVAYIITSGTTGKPKGVVIE 4771
Qy 67 ----- 66
Db 4772 HRSYANVAFANKDEYHLDSFPVRLQNASFADFVSTGDFARALLTGQGVICPNGVKMDP 4831
Qy 67 ----- 66
Db 4832 ASLYETIRRHEITIFEATPALIMPLMHYVENELDSOMKLLILGADSCPAEDPKTLAR 4891
Qy 67 ----- 70
Db 4892 FGQKWRINSYGVTEACIDTSYBETDVTAIRSGTVPFGKPLPNMTWYVVDHNLQPVG 4951
Qy 71 ----- 70
Db 4952 VVGELCIGGAGVARGYLNRPETBEKFPNPFAPGERLYRTGDLAKWRADGNVEFLGRND 5011
Qy 71 ----- 72
Db 5012 HQVKIRGVRIELGIETQLRKLDGITEAVVVVAREDRGOEKELCAYVVADHKLDTAELRAN 5071
Qy 73 LTSSL-----LSVCG----- 82
Db 5072 LKELPQNMIPAYFVTLDALPLTANGKVDRRSLPAPDVTMLRTTEYVAPRSVWEARLAQV 5131
```

```
Qy 83 WSOTIN----- 88
Db 5132 WEQVLNVPOVGALDDFFALGGHSLRAMRWLSSMHNEYQVDIPRILRILPEKPTIOELAAPIE 5191
Qy 89 ----- 88
Db 5192 ETAKGNVFSIEPVQKQAYYPVSSAOKRMVILDOFEGVGISYNMPTMLIEGKLERTRVEA 5251
Qy 89 ----- 88
Db 5252 AFORLIARHESLRTSFAVNGEPVONIHEDPFALAYSEVTEOEARELVSLSVQFPDLEV 5311
Qy 89 ----- 88
Db 5312 APLIRVSLLKIGEDRYVLFTDMHHSISDVSSGILLAEWVOLYQGDVLPBLRIQYKDFAV 5371
Qy 89 ----- 95
Db 5372 WQOEFSSAAAFHKQEAYNLOTFPADDIPVLNLPDTPTRPSTOSFAGDOCTIGACKALTEGL 5431
Qy 96 ----- 100
Db 5432 HOLAQATGTTLYMVLAAAYNVLLAKYAGQEDIIIVGTPIITGRSHADLEPIVGMFVNTLAMR 5491
Qy 101 ----- 100
Db 5492 NKQREKTFSEFLOEVKQNALDAYGHQDYPFEELVEKLAIRDLSNPLPDTVTFQNST 5551
Qy 101 ----- 107
Db 5552 EEVMTLPBCTLAPFMTDETGOHAKFDLTTSATEEREEMTIGVYSLSLTRETMEFRSRH 5611
Qy 108 ----- 117
Db 5612 FLTIAASIVQNPHIRLGEIDMLLPEEKQIILAGFNDTAVSYALDKTLHOLFEEQVOKTPD 5671
Qy 118 ----- 126
Db 5672 QAALLFSEOSLTYSSELNERANRLARVLRAKGVDPRLVAIMAERSPEMWIGILGILKAGG 5731
Qy 127 A----- 127
Db 5732 AYVPDPGYPQERIQYLLLEDSSNALLLSQAHLLPQLLAQVSSSELPECLDLNAELDAGLSGS 5791
Qy 128 ----- 127
Db 5792 NLPVAVNQPTDLAYVIYTSGTTGKPKGVMIPHOGIVNCLQWRDEYGGPSDKALQVPSFA 5851
Qy 128 ----- 129
Db 5852 FDGFVASLPAPLLGGATCVLPQBAAKADPVALKKMAATEVTHYGVPSLQFQAILDCSTT 5911
Qy 130 ----- 137
Db 5912 TDFNQRVCVTLGKELPVQLVQKTKKHPAIEINNEYGPTENSVVTTISRSIAGQAITI 5971
Qy 138 ----- 137
Db 5972 GRPLANVQVYVDEQHHLOPIGVVGBELCIGGACLAGLYLNKLPBELTAEKFPANPRPOERM 6031
Qy 138 ----- 137
Db 6032 YKTGDLVKWRTDGTIBYIGRADEQVKRGYRIEIGIESAVLAYOGIDQAVVVVARDDDAT 6091
Qy 138 ----- 137
Db 6092 AGSYLCAYFVAATAVSVGLRSHLAKELPAYMIPSYFVELDQPLSANGKVDRLKALPKPO 6151
Qy 138 ----- 137
Db 6152 QSDATTREYVAPRNATEQQLAAIWQEVLGVEPIGITDQFPELGHSLKATLLIAKVVEYM 6211
```


Qy 138 ----- 137
 Db 6212 QIELPLNLFQPTIEKVADFIHKRFESRYGTAILLNOETARNVFCFTPIGAQSVYQK 6271
 Qy 138 -----EDT 140
 Db 6272 LAAETOGVSLYSFDFIQQDNRMEQYIAAITAIDPSGPTLMGYSGGNLAFEVAKELEER 6331
 Qy 141 GFDLGV 147
 Db 6332 GV-GVT 6336
 RESULT 12
 ID ACVS NOCLA STANDARD; * PRT; 3649 AA.
 AC P27743.
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Delta-(L-alpha-aminoadipyl)-L-cysteiny-D-valine synthetase
 DE (EC 6.-.-.-) (ACV synthetase) (ACVS).
 GN PCBAB.
 OS Nocardia lactamurans.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae;
 OC Amycolatopsis.
 OX NCBI_TaxId=1913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VAR LC 411;
 RA MEDLINE=92065808; PubMed=1956290;
 RX Coque J.J.R., Martin J.F., Calzada J.G., Liras P.;
 RT "The cephamycin biosynthetic genes pcbAB, encoding a large
 multidomain peptide synthetase, and pcbC of Nocardia lactamurans are
 clustered together in an organization different from the same genes
 in Acromonium chrysogenum and Penicillium chrysogenum.";
 RT Mol. Microbiol. 5:1125-1133(1991).
 RL
 CC -!- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE
 ACV ARE ACTIVATED AS AMINOACYL-ADENYLATES WITH PEPTIDE BONDS
 FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER
 INTERMEDIATES.
 CC -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHETHEINES.
 CC -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND
 CEPHALOSPORIN.
 CC -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 FAMILY.
 CC -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X57310; CAB40561.1;
 DR PIR; S18268; S18268.
 DR HSP; P14687; LAMU.
 DR InterPro; IPR000873; AMP-bind.
 DR InterPro; IPR001242; Condensatn.
 DR InterPro; IPR003880; Ppantne attach.
 DR InterPro; IPR000379; Ser extra site.
 DR InterPro; IPR001031; Thioesterase.
 DR Pfam; PF00501; AMP-binding; 3.
 DR Pfam; PF00550; pp-binding; 3.
 DR Pfam; PF00668; Condensation; 3.
 DR Pfam; PF00975; Thioesterase; 1.
 DR PRINTS; PR00154; AMPBINDING.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 2.
 DR PROSITE; PS00455; AMP BINDING; 1.
 DR PROSITE; PS00075; ACP_DOMAIN; 3.

Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
 Repeat; Phosphopantetheine.
 FT REPEAT 401 861 DOMAIN 1 (ADIPATE-ACTIVATING).
 FT REPEAT 1014 1937 DOMAIN 2 (CYSTEINE-ACTIVATING).
 FT REPEAT 2079 2985 DOMAIN 3 (VALINE-ACTIVATING).
 FT DOMAIN 788 857 ACYL CARRIER (ACP) 1.
 FT DOMAIN 1864 1933 ACYL CARRIER (ACP) 2.
 FT DOMAIN 2910 2981 ACYL CARRIER (ACP) 3.
 FT BINDING 820 820 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT BINDING 1896 1896 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT BINDING 2944 2944 PHOSPHOPANTHETHEINE (BY SIMILARITY).
 FT ACT SITE 3502 3502 THIOESTERASE (BY SIMILARITY).
 SQ SEQUENCE 3649 AA; 404079 MW; 6FD095704F85986B CRC64;
 Query Match 27.0%; Score 209; DB 1; Length 3649;
 Best Local Similarity 2.6%; Pred. No. 27;
 Matches 81; Conservative 23; Mismatches 43; Indels 2965; Gaps 24;
 Qy 1 AAG-----GIL----- 6
 Db 180 AAGCLHRIAYAGEFFDKIIAGVLDVAREVLGQFGRPEQLVADILVSABOELQHQW 239
 Qy 7 ----- 6
 Db 240 NCTGEPEDKRLNELFEDVVRAPDREAVVCCDVRLTYREVRERANQFAHMLIOGPVRV 299
 Qy 7 -----H 7
 Db 300 RPGALIGLYDKSLGVVATFGIMKSGAAYVPIDPAYPAERIRFLVGTGLSGIVTNRHH 359
 Qy 8 LELL-----VAVGPD----- 17
 Db 360 AERLEVLGDEHASVHVIEVAWAGPHPEQARENPGALSRRDAYVYVYTGTTGVPKG 419
 Qy 18 -----VFOAHOEDT----- 26
 Db 420 VPKHYSVNSITDLSERYDMRRPOTERVALFASVVFPHLRLQTLIALINEQTLVIVPD 479
 Qy 27 -----ERYVLTLN----- 35
 Db 480 VRLDPDLFPYIERHGVYLNATSGVLQFHLRRCASLKRLLVGEELTASGLRQREKF 539
 Qy 36 -----TGA----- 38
 Db 540 AGRVVEYAFTEAAFTAVKEFGVGTERRDRSIGRPLRNKVMYVLSOGLKOLPGAIGE 599
 Qy 39 ----- 38
 Db 600 LYIGCGVAPGLNRRDILTAFRTANPQTTEEKARGNRLYRTGDLARVLLNGEVEPM 659
 Qy 39 ----- 38
 Db 660 GRADFQKLNGVRVPEGIEAQATEFPGVKCVVAKENATGDRHLVGYYLVEDGAETAE 719
 Qy 39 ----- 38
 Db 720 ADLIAFLEQLIRIMVARMVRLTSPVNVNGKVDWRALPDVSLHPAPANMNGALLAID 779
 Qy 39 ----- 38
 Db 780 GSNAPLLAITEOLRAIWSEVLGVQPNRIGERDDPFRLLGQSISCILLIARVQRSLSLG 839
 Qy 39 ----- 38
 Db 840 VEDFALRTLDAAGHLESQGHAEPEVVAEEVTTGSEPVRLANGLQQGLLYHLLKTAGG 899
 Qy 39 -----ELLRD-----PSL----- 46
 Db 900 DDAVVQSVHYHAPIRPELMKDAQAAQYTPALRLRFDWAEPEVQIVDNDKPPDMRF 959
 Qy 47 -----GAQFRVHLVKM----- 57
 Db 960 VDLSATADDAEQEARVRELQERDRTEPYDLAGGLRFLRYLIKQREDLFLSIFSCHHIL 1019

```
QY 58 ----- 57
Db 1020 GWSLPVLHDEVHRNYLALRAGQPIESVDNAYVAAQRYWEAHRNDHAAHYVVEQLGRIDER 1079
QY 58 ----- 57
Db 1080 GDFAGLLNEKSRYSRLGDYDHVQRHRTRKLYLGADLTGALKAGCAADQVTLHSLVQFVW 1139
QY 58 ----- 57
Db 1140 HKVLHAIGGGNTTVGTIVSGRNLPGDIENSAGLFINTLPLIVDHQQQAGQNVAAEVRD 1199
QY 58 ----- 57
Db 1200 IQAAVNTWNSKISVELGRQSGEMKRRLFDTLVLENYPRLLDEEBELAHQEARFEKAY 1259
QY 58 ----- 57
Db 1260 DADKVDYPIAVAREEGDELTVTLWYAGELFDEDTIDTLDDVARTLFRQVTEDIARPVE 1319
QY 58 ----- 57
Db 1320 LDLISPMRARFDSWNETAEEFPADKTLHAVFEEMAERWPDEIAVYVRENRLTYRELNER 1379
QY 58 ----- 57
Db 1380 ANRLAHLRSVVELRPDDLVALVLDKSELMITAIIAAWKTGAAVVPIDSGYPDDRISFML 1439
QY 58 ----- 57
Db 1440 SDTAARVVVWNEIHSRDLRLSAETCTPVLEITELLHDDQPAVNPVTETSTDLAYAYTS 1499
QY 58 ----- 73
Db 1500 GTTGKPKAVLVEHRGVNQLVSLAKLFOLDKAHRDEALLSFSNYIFDHFEQMTDALLNG 1559
QY 74 ----- 79
Db 1560 QKLVVLDSDMRDQRLCRYNDEQVTVLSGTPSVLSLYDYSSATSLTRIDAIGEDTEP 1619
QY 80 ----- 79
Db 1620 VFAKIRCTFPGLIINGYPTISITSHKRPYPDPVHRVKNKSGIPPVANTKCHVLNKAMKP 1679
QY 80 ----- 79
Db 1680 VPVGIGELYIGGIVTRGYLNREDLTADRFVENPFOAEBERRLGENGRLYKTGDLVRWL 1739
QY 80 ----- 79
Db 1740 PNGEVEYLGRTDLQVKIRGORVELGEVEAALSSYPGVVRSLVVAREHAVGOKYLVGFVVG 1799
QY 80 ----- 79
Db 1800 EQEFDEQDLKQWMRKLPESVVPARVLRTIDIVTPPSGKLDARRLPETDFCAGEGEYVA 1859
QY 80 ----- 88
Db 1860 PVSEFELKLCGIWAQVLEIAPDRIGVHDDFFALGSDSIRAMALAQAITTFGQGLGVATV 1919
QY 89 ----- 88
Db 1920 LOHTTLAAQAEHIQAAALEHTAWPTTAVEHPVPSLAQERLLLFIDDFEGGTAAVNPVFV 1979
QY 89 ----- PEDD ----- 92
Db 1980 LRLPAHTRALPGALGTLVRRHPALRTLKTDQGRQYRPIPADDDVRLVLEVPSTTVDSRA 2039
QY 93 ----- 92
Db 2040 ELDEVLTIRAGYVFRLHBEELPIRAEAFDHGDEIYLSVVVHHSCFDGWSWDIFRRELAALL 2099
QY 93 ----- 92
Db 2100 DGYPEADLGLRGCTGYEFQVWQRYLTGKRLAALTEFWTGCALGPFETIALPDHPRPRF 2159
QY 93 ----- 92
Db 2160 DYRGRELEFELDERTTREALRELARTARVSLYSVLLGAWCLMLNMYTQCHDLVVGTPSANR 2219
QY 93 ----- 104
Db 2220 GRPEFORAVGFFPANLLALRVRVDPAAATLPAYVRSVGEAVVAAQVHGCELPFEQLVKELKV 2278
QY 105 ----- 110
Db 2279 EKDPSRHPILQLNFTLQNVSDHTSALTGYQPDGGMTTTFKDLTSATMTETATGLAGNLTY 2338
QY 111 ----- 116
Db 2339 AASLFDSDTSASGFIATFKHVLAEFASAAAQTPIAQLTALDEPQOALPDATRRARRPGGP 2398
QY 117 ----- 116
Db 2399 GRCTRLEFEVAATWPDRAVAVHGDVRLTYRELNERANRLAHLRSVAEPRADELIALVLD 2458
QY 117 ----- 123
Db 2459 KSELTILVAILAVNKAGAAVMPIDPSYPDDRIAFLMSDTGAKVLAGEAHGSRVRLTSGD 2518
QY 124 ----- 123
Db 2519 VLDLEQLDLTGEPAPNPVTETSTELAYAIYTSGTTGPKKAVLVSHSGVDSFRAQLSGRY 2578
QY 124 ----- 126
Db 2579 FGSFDESABEAVFLANYVDFSVBQLALSVLGGHKLPPPPSAADDPAFYELANREGLSY 2638
QY 127 ----- 126
Db 2639 LSGTPTQVERFDLARLSHLRCVLVAGEAFQPHFKMRGEFAGPILNAYGTTTETTVNTV 2698
QY 127 ----- 126
Db 2699 HRPEPGDAYRNTLGAPLGNTRLVYLGDMKLLPTGAVGELYLAGDCVTEGYLHRPELTRE 2758
QY 127 ----- A 127
Db 2759 RFLPNPFAESGRFPMTYRTGDUVRGPDGELQYLGNDQAQVINGLRIEPEGEVEAALAG 2818
QY 128 CSPTWMSCL1 ----- 136
Db 2819 CSGVROCAVVAGADPOAPERKRLVGYVYLPPEGAADVDEADLFAALRAQLMPSVPSLLVRL 2878
QY 137 ----- 136
Db 2879 DRBLPMTITGKLDVADLPSADSPKRAAYAAPDRVRLEARLCHLWSAQLPGGTGVIDDDFF 2938
QY 137 ----- 136
Db 2939 RCGGDSISALHULASOVOREIERKVSVKYLFDPHTVRSFVDNVLISGLAESGDDPEQOGL 2998
QY 137 ----- 136
Db 2999 TGECPLPIQEWFFAKPLADHRHRWNHFAIRTPPLDPGELRTALDRLVEHHDAFRLRPE 3058
QY 137 ----- 136
Db 3059 SGGEVYAEADAPITLHELDVRLGADADLRLQRLVDWORTFDLANGPTCAAYLHGHFDGCTA 3118
QY 137 ----- 136
Db 3119 RVMPALHHLVVDTVSWHILAOQLEILYNGDGLGNKTSYRQWAAQVRDYTPAEGEREFWA 3178
QY 137 --TBD-----TGF--- 142
```

Db 3179 ETTRDMESAELLAQTEGTRRRREFALTAPDTRTLAESPWAYDTEVNDLLLTATGFAIR 3238
QV 143 -----DLQVTI 148
Db 3239 SITROATNHLTVEGHGRELFCGAPDVRTVGFTTWHFPAVEVDPGDLGRSV 3290
RESULT 13
FAT_DROME STANDARD; PRT: 5147 AA.
AC P33450; Q5VQX5;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cadherin-related tumor suppressor precursor (Fat protein).
GN FT OR CG3352.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92069752; PubMed=1959133;
RA Mahoney P.A., Weber U., Onofrechuk P., Biessmann H., Bryant P.J.,
RA Goodman C.S.;
RT "The fat tumor suppressor gene in Drosophila encodes a novel member
RT of the cadherin gene superfamily.";
RL Cell 67:853-868(1991).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laek P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milghina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphay L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN. ACTS AS A

CC CC
CC TUMOR SUPPRESSOR. REQUIRED FOR CORRECT MORPHOGENESIS.
CC -!- SIMILARITY: CONTAINS 34 CADHERIN DOMAINS.
CC -!- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M80537; AAA28530.1; --
CC EMBL; AE003577; AAF51036.1; --
CC PIR; A41087; IJFFTM.
CC HSP; P00740; 1EDM.
CC FlyBase; FBgn001075; ft.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR000742; EGF 2.
CC InterPro; IPR001791; Laminin_G.
CC Pfam; PF00008; EGF; 4.
CC Pfam; PF00028; cadherin; 34.
CC Pfam; PF00054; laminin_G; 2.
CC PRINTS; PR00205; CADHERIN.
CC SMART; SM00112; CA; 34.
CC SMART; SM00181; EGF; 4.
CC SMART; SM00282; LamG; 2.
CC PROSITE; PS00232; CADHERIN_1; 22.
CC PROSITE; PS0268; CADHERIN_2; 34.
CC PROSITE; PS00022; EGF 1; 4.
CC PROSITE; PS01186; EGF 2; 2.
CC PROSITE; PS50025; LAM_G DOMAIN; 2.
CC Cell adhesion. Signal; Transmembrane; Glycoprotein; Calcium-binding;
CC Repeat; EGF-like domain.
CC SIGNAL 1 35
CC CHAIN 36 5147
CC DOMAIN 36 4583
CC TRANSMEM 4584 4609
CC DOMAIN 4610 5147
CC DOMAIN 36 156
CC DOMAIN 157 270
CC DOMAIN 271 382
CC DOMAIN 383 494
CC DOMAIN 495 599
CC DOMAIN 600 708
CC DOMAIN 709 820
CC DOMAIN 821 942
CC DOMAIN 943 1049
CC DOMAIN 1050 1153
CC DOMAIN 1154 1278
CC DOMAIN 1279 1384
CC DOMAIN 1385 1489
CC DOMAIN 1490 1601
CC DOMAIN 1602 1713
CC DOMAIN 1714 1823
CC DOMAIN 1824 1922
CC DOMAIN 1923 2027
CC DOMAIN 2028 2167
CC DOMAIN 2168 2278
CC DOMAIN 2279 2385
CC DOMAIN 2386 2491
CC DOMAIN 2492 2596
CC DOMAIN 2597 2703
CC DOMAIN 2704 2810
CC DOMAIN 2811 2913
CC DOMAIN 2914 3013
CC DOMAIN 3014 3124
CC DOMAIN 3125 3229
CC DOMAIN 3230 3334
CC DOMAIN 3335 3439
CC DOMAIN 3440 3545
CC
CC POTENTIAL.
CC CADHERIN-RELATED TUMOR SUPPRESSOR.
CC EXTRACELLULAR (POTENTIAL).
CC POTENTIAL.
CC CYTOPLASMIC (POTENTIAL).
CC CADHERIN 1.
CC CADHERIN 2.
CC CADHERIN 3.
CC CADHERIN 4.
CC CADHERIN 5.
CC CADHERIN 6.
CC CADHERIN 7.
CC CADHERIN 8.
CC CADHERIN 9.
CC CADHERIN 10.
CC CADHERIN 11.
CC CADHERIN 12.
CC CADHERIN 13.
CC CADHERIN 14.
CC CADHERIN 15.
CC CADHERIN 16.
CC CADHERIN 17.
CC CADHERIN 18.
CC CADHERIN 19.
CC CADHERIN 20.
CC CADHERIN 21.
CC CADHERIN 22.
CC CADHERIN 23.
CC CADHERIN 24.
CC CADHERIN 25.
CC CADHERIN 26.
CC CADHERIN 27.
CC CADHERIN 28.
CC CADHERIN 29.
CC CADHERIN 30.
CC CADHERIN 31.
CC CADHERIN 32.

FT	DOMAIN	3546	3651	CADHERIN 33.
FT	DOMAIN	3652	3756	CADHERIN 34.
FT	DOMAIN	3950	4011	EGF-LIKE 1.
FT	DOMAIN	4013	4049	EGF-LIKE 2.
FT	DOMAIN	4052	4090	EGF-LIKE 3.
FT	DOMAIN	4092	4128	EGF-LIKE 4.
FT	DOMAIN	4129	4320	LAMININ G-LIKE 1.
FT	DOMAIN	4321	4362	EGF-LIKE 5.
FT	DOMAIN	4402	4569	LAMININ G-LIKE 2.
FT	DISULFID	3954	3966	BY SIMILARITY.
FT	DISULFID	3960	3999	BY SIMILARITY.
FT	DISULFID	4001	4010	BY SIMILARITY.
FT	DISULFID	4017	4028	BY SIMILARITY.
FT	DISULFID	4022	4037	BY SIMILARITY.
FT	DISULFID	4039	4048	BY SIMILARITY.
FT	DISULFID	4056	4067	BY SIMILARITY.
FT	DISULFID	4061	4078	BY SIMILARITY.
FT	DISULFID	4080	4089	BY SIMILARITY.
FT	DISULFID	4096	4107	BY SIMILARITY.
FT	DISULFID	4101	4116	BY SIMILARITY.
FT	DISULFID	4118	4127	BY SIMILARITY.
FT	DISULFID	4325	4341	BY SIMILARITY.
FT	DISULFID	4334	4350	BY SIMILARITY.
FT	DISULFID	4352	4361	BY SIMILARITY.
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	257	257	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	276	276	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	280	280	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	402	402	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	461	461	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	605	605	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	631	631	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1155	1155	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1367	1367	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1458	1458	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1751	1751	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1831	1831	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1880	1880	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2080	2080	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2171	2171	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2247	2247	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2290	2290	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2437	2437	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2581	2581	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2799	2799	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2920	2920	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2946	2946	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2967	2967	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3167	3167	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3303	3303	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3386	3386	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3389	3389	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3525	3525	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3852	3852	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3865	3865	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3905	3905	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4306	4306	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4414	4414	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4471	4471	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4487	4487	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4539	4539	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	4550	4550	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARIANT	1229	1229	S -> G.
FT	VARIANT	1233	1233	G -> S.

Query Match 27.0%; Score 209; DB 1; Length 5147;
Best Local Similarity 1.8%; Pred. No. 76;
Matches 86; Conservative 23; Mismatches 36; Indels 4763; Gaps 29;

Qy	2	AGGI-----LHLE-----	9
Db	198	AGVNDKFRVLTTPNSGDTSYLHLETTGNLDRESRGYQLNISARDGSPRFGYQVN	257

Qy	10	-----	9
Db	258	VTILOVNDNPPIFDHSDYNSLNETALPGTPVVTVMASNDLDGNSKITYYLAETEHQFT	317
Qy	10	-----	9
Db	318	VNPETGVISTTERVNCPOQTNVKSASOKSCVTFVFARDHGSRODGRTYVTVNLDTND	377
Qy	10	-----	9
Db	378	HDPISFRFFPDGGKVATVDENAVNGTVVAAVAKDSGLNGRTSVRIVSGNELGHFRL	437
Qy	10	-----	9
Db	438	EEAADLHIVRVNGVLDREIEIGYKNTVVMDOQTPTATTAAHLIIDVNDVNDHEPVFEKS	497
Qy	10	-----LLVAVGP- 16	
Db	498	EYSAVLSLAPTGSFVASITATDEDTGVNAQVHYDILSGNELKWFMDPLTGLIVTTGPL	557
Qy	17	-----	16
Db	558	DREIRDTVELSISARDGGPNPKFAYTQLVIIIDENDEAPQFSOREQNVTLGEDAPQTI	617
Qy	17	-----	16
Db	618	VALMTATDHDQGTNGSVTFALAPSVERLYPLQFALDALTGQLTTRRPLDREKMSQYEISV	677
Qy	17	-----	16
Db	678	IARDQAGTPQSATATVNLVADVNDNDPQFYPRHYIYSLADDDDDIKLKKEVEKERILL	737
Qy	17	-----	16
Db	738	HVTASDKDDGDNALIEYRLSESGEGLFQLDARSGAISLRGDAPASHMHKPHYKLLVSARD	797
Qy	17	-----	16
Db	798	AGORRQQDAIVEIVLKSLEMCQAAGGYEQFQWEDHEQORNSQPNREVGIVQVKS	857
Qy	17	-----	16
Db	858	TNGKANSHEYDIIQGDRAQNFRTDTRSGRITARTPLDREEQANYRLTILASSSSSSAA	917
Qy	17	-----	16
Db	918	ASSVSYGCIVNTAII DLNDNAPVFDRESEPTISLPENAVGQEIYLSRVDRDAGVN	977
Qy	17	-----	16
Db	978	SRISYSLTNPNQOFRIGPVTVGLYQRPPIRAEPGSLIHVELMATDAGSPPLSKLSLV	1037
Qy	17	-----	16
Db	1038	LIADVNDTPVFOHTSVETSLPETTKVNTFRFFALAATDIDLGDNGRISYIEIGTERMF	1097
Qy	17	-----	16
Db	1098	GVFPDGYLFVRAPLDREERYALTVCSDRAGQPSRSSVVPVVIHVIDENDNAPQFTNST	1157
Qy	17	-----	16
Db	1158	FTSIPENAPADTFVGKLTAVDRDIGNAELSFLLSSQOTDFTIDTRNGFIKTLRPFDR	1217
Qy	17	-----	16
Db	1218	ALVKVSRNAESGDSGLRSGMAGNYMLLEATVSDNGIPRLQDKVKVIVTDVNDNAPE	1277
Qy	17	-----	16
Db	1278	FLRAPHVTISEGASEGTHITHVFTQDAEGLNGDVVYSLAKNEAQFNLDATGOLS	1337
Qy	17	-----	16

```
Db 1338 GRRLDRESQEIHHLIIVAKDAALKHPLSSNASITIVILDENDNAPFTQSSSEVSLETS 1397
Qy 17 -----16
Db 1398 PTGTELMFRASDADQGVNSQVVFISAGNRRDTHIDSITGSLYLHKPLDYEDITSYTL 1457
Qy 17 -----16
Db 1458 NITASDCGTPSLSTTVLYNVLVDDNDNPPFPSTAIVRQIKEGIPLKPTIVTVTADDDP 1517
Qy 17 -----16
Db 1518 SGLNGKSVYAISKQBPOLPQGRHFGINTETGVHTLREIDRESIDTFLTVVATDRAQPS 1577
Qy 17 -----16
Db 1578 ERQSTKLVTVIVEDINDNPVFMVNMNAAILPPKFSTSGSSTAIVMQVHAKDADSSNG 1637
Qy 17 -----16
Db 1638 LVTYEIVSGPQELFKLQRTGIIITTPGPQKQEVRYQLTKSTDEAVQSERRSSEVIT 1697
Qy 17 -----16
Db 1698 IITPGSGSESVPOFEQORSKLSGVYENEPIGTSILTVAHLASAEIEYFVTVNVTATGS 1757
Qy 17 -----16
Db 1758 RGQVRLFDIDAKGILSTAAELDREAGPEEYEVYAIAGQGQPTSRKRVTVLDKN 1817
Qy 17 -----16
Db 1818 DSPQFLOTFFVYVNSEDLQIGHTISTLRAHDPDTLGSVTFLLMDGHGKFLLEPSTCKL 1877
Qy 17 -----16
Db 1878 ILNDLTRKTSKYLIRVSDGOVTEAYATIOVSDTNDNPPLFEDTVYSPDIPENAAQR 1937
Qy 17 -----DVFO-----20
Db 1938 GYQVGQIWARDADLQNAQLSYGVVSDWANDVFSINPOTGMLTLTLARLDYEEVQHYILIV 1997
Qy 21 -----20
Db 1998 QAQDNGQPSLSTTIIVYCNVLDLNDNAPIFDPMSYSSEVFENVPIATEWTVYSAKDIDSG 2057
Qy 21 -----20
Db 2058 NNGLIYSITAGDVDFGIDSGNGTIRTRNLDREHRSYTLTVTARDCADEFASFSELE 2117
Qy 21 -----AHQ-----23
Db 2118 ETQLKLYRSRKYQOTROEFLAHQKQRLSSTVKVITLIKVDNDEVVPFISANETAIME 2177
Qy 24 -----23
Db 2178 NVAINTVIAKAVDNDEGRNGYIDYLMKEARDEDMQSDPLPFSLNPTDGLRVVDALD 2237
Qy 24 -----23
Db 2238 RELRSSYLLNITARDGEPPQSTESOLLIRILDENDNSPVDFPKOYSASVAENASIGAMV 2297
Qy 24 -----EDT-----ER-----28
Db 2298 LQVSATDVEGANGRIRYSIVLGDQNHDFSISEDGTGVRVAKNLNYERSYSLTVRAED 2357
Qy 29 -----YVLT-----32
Db 2358 CALENPAGDTAELTINILDNDNRPTFLDSPYLARVMENTVPPNGGYVLTVNAYDADTPP 2417
Qy 33 -----32
Db 2418 LNSOVRYFLKEGSDSLFRINASSGDIALKPLDREQOSEYTLTLVAMDTGSPPLTGTGIV 2477
Qy 33 -----NLNIGAEEL-----RDPSLGAOFR-----51
Db 2478 RVEQDINDNDPVPFELQSYHATVRENLPSTGTHVLTPRATDKDEGLNAKRLNGLGEHMR 2537
Qy 52 -----VHLVKM-----57
Db 2538 FHIDSETGEISTATTLTREETSVVHLTLMAQDSSITEPRASSVNLTISVSDVNONIPKFD 2597
Qy 58 -----57
Db 2598 STTYNAVPERISKGEFVFGARALDLDGGENAVVHYTISGRDQHYFDINTKTGVVSTKLE 2657
Qy 58 -----VILTEPE-----64
Db 2658 LKTKTSHDDLTYTIVISAMDQEOSLSSKAELTVILRPPPELFPFTFAYMANSHFAMSBDV 2717
Qy 65 -----64
Db 2718 RPKXNITKVSATSPKGLVGKIRYAIAGGIMGDSLRVDPNPSGLLSVGQDGLDYELTHLYE 2777
Qy 65 -----64
Db 2778 IWIEAADGDTPLSRSVTLITLVNTDANDNAPVMEQLIYNAEVLSEESPPQLIAVVKASDR 2837
Qy 65 -----GAP-----67
Db 2838 DSGDNGNVIYRLQNDPDTTEITSEGEIYTRMLRDREBIGDYAFVVEAVDQGVPHMTGTA 2897
Qy 68 -----NITAN-----LTSS-----76
Db 2898 SVLHLLDKNDNPKPKFTRFLSNVTENAEIGSFVIRVTSSDLDIGANANASYSPSENPE 2957
Qy 77 -----76
Db 2958 KFRIEPOSGNITVAGHLDRQODEYILKVNASDGAWEAETPITITIQDNDNAPEFHSF 3017
Qy 77 -----76
Db 3018 YSFPPELQQSIALVGQIIATDRDKQGNPSVISYSLOQSPMFSIDPATGEVFSKAVRF 3077
Qy 77 -----LLSVC-----81
Db 3078 KHSOYVRSPENMYALTVLATDNGKPLYSECLVNIINIVDAHNNPKPEQAEYLAPLPODA 3137
Qy 82 -----GWSOTINP-----89
Db 3138 VRGQIRVVRHANDKQDLGTNEMDYSMTFNLSISFVGRHGWITLVKPIQVPPNTRYEL 3197
Qy 90 -----BDD-----92
Db 3198 VVRATDRGVPQSDETRVIIVVTGENMDTPRFSVNSYQVIVPENEPVGSTILTGVATDDD 3257
Qy 93 TDPG-----96
Db 3258 TGPNGMLRYSISGGNERQDFSVDBERTGGIVIQOQLDYDLIOEYHLNITVQDLGYHPLSSV 3317
Qy 97 -----96
Db 3318 AMLTITLTDVNDNPPVFNHKEYHCYIPENKPVGTFFQAHAAKDQSPKNAIHYAFLPSG 3377
Qy 97 -----96
Db 3378 PDRHFFIMNQSGNITSSAVSFDYBERRIYTLQIKAKNPSSSMESYANLYVHVLGVNEFPY 3437
Qy 97 -----HADL-----VLY 103
Db 3438 QFLOPVFHFVDFVSETSAVGTRVGAVQATDKSGEDGRVYLLVGSNDKGFRIIDTNTGLTY 3497
Qy 104 ITR-----106
Db 3498 VARHLDRQNRVVLTVNAKNYGSIRGNDTDEAQVVIISIQDGNDDPEFIKHYTSTISEA 3557
```

```
QY 107 ----- 106
Db 3558 APVGTKVTTKAKDKVTONNQFSYIINGNLKQSFKIDVQTGEISTASRLDRETSY 3617
QY 107 ----- 106
Db 3618 NLVIGAITDGLPQTGSATVHIELEDVNDNGPTFTPEGLNGYISENEPAGTSIMTLIASD 3677
QY 107 ----- 111
Db 3678 PDLPRNGPPTYQLIGGKHKSWLSVDRNSGVVRSTTSFDREMTPILEAIIIEVEDSGPKPK 3737
QY 112 ----- 111
Db 3738 KSHLLTITVLDQNDNPSTRSLHTIAVSLFNGDLPNSVKLADVRPNIDIVGDYRCRLQK 3797
QY 112 ----- 111
Db 3798 NPAQSOLQLAI PRACDLITTSHTPIASVFSYTGNDGKHCDVSSKVSVAFOFNNETLAN 3857
QY 112 ----- 111
Db 3858 SVSIVMRNMTAYHFLANHYRPILEMIKSRMSNEDEVILYSLLEGGSGNSTNLQLMAVRL 3917
QY 112 ----- 111
Db 3918 AKTSYQPKYLIBRLREKRSFAFSELLQKEVIVGVEPCSEPDVCGNGVCSATMRLDDAHS 3977
QY 112 ----- 111
Db 3978 FVIQDSPALVLSGRVVRVHDYSOCCTSGFGSQCRSRQDPCLPNPCHSQVQCRRGLSGDFQC 4037
QY 112 ----- 114
Db 4038 MCPANRDGKHEKERSDVCYSKPCRNCGSCQSPDGSSEYFCLCRPGFRGNQCESVSDSCR 4097
QY 115 ----- 114
Db 4098 PNPLHGLCLVSLKPGYKCNCTPGRYGRHCFERSYGFQPLSYMTFPALDVTNTDISIVFA 4157
QY 115 ----- 114
Db 4158 TTKPNSLLLYNMQSGGRSDFLAIELVHORAYFSSGGARTALSTVIAGNLADGWHKV 4217
QY 115 ----- 122
Db 4218 TATRNGRVMSLSVAKCADSGDVCTECLPGDBSCYADEVGPVGTLENFNKQPLMIGLSSAD 4277
QY 123 ----- 131
Db 4278 PILERPQVHSDDLVGLCHSVHIGGRALNLSLPLQOKGILAGCNROACQPALAAERCGGF 4337
QY 132 ----- 133
Db 4338 AGOCIDRWSSSLCQCGHLOSPDCSDSLEPITLCEGAFVFRISEYRRMQLLDNLNYSK 4397
QY 134 ----- 133
Db 4398 SANLDNQMRERRAVSNFSTASQIYEAPKMSMLFRITYKQGGQILYAATNQMTLSLRE 4457
QY 134 ----- 133
Db 4458 GRLVYYSKQHLINMTVQETSTLNDGKWNHNSLFSERSLRLIVDGRQVGDELDIAGVHD 4517
QY 134 ----- 136
Db 4518 FLDPYLTILNVGGEAFVGCCLANVTVNNELQPLNGSGSIFPEVRYHKGIESGCRGIGQDA 4577
QY 137 ----- 136
Db 4578 AQVADPLSIGFTLVIVFFVLVAILGSYVIYRPRGKQEKIGSLSCGVPFGFKIHGPGVP 4637
```

```
QY 137 ----- 136
Db 4638 TQSQVDHVLVRLNHPSEAPSPVAGDHMRPVGSHLVLGPPELLTKKFKPTAEMPQPO 4697
QY 137 ----- 136
Db 4698 QOQRQRPDIIRERESPLIREDHHLPIPLHPLPLEHASSVDMGSEYPEHYDLENASSIAP 4757
QY 137 ----- 136
Db 4758 SDIDIVYHYKYGREAAGURKYKASVPPVSAYTHHKQNSGSOQOQOQOQHRHTAPFVTRNOG 4817
QY 137 ----- 136
Db 4818 GQPPPTTSASRTHOSTPLARLSPSELSSQOPRILTLHDISGKPLQALLATTSSSGGV 4877
QY 137 ----- 136
Db 4878 GKDVSNSERSLNSPMSQLSQSSASROKPGVPOQAQOOTSMLGTABEIERLNGRPRT 4937
QY 137 ----- 143
Db 4938 CSLISTDAVSSSEAPRVSSSALHMSLGGVDHAHSTSTDSGNDSTFCSEIEYDNNSL 4997
QY 144 ----- 143
Db 4998 SCDGKYSTSKSLDGRSPVSRALSGGETSRNPPTTVTKTPIPPHAYDGFESSFRGSLST 5057
QY 144 ----- LGV 146
Db 5058 LVASDDDIANHLGSIYRKANGAAPSATTLGWYLLNMGPSYENLMGV 5105
```

RESULT 14
PKD1_HUMAN
ID PKD1_HUMAN STANDARD: PRT: 4303 AA.
AC P98161;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Polycystin precursor (Autosomal dominant polycystic kidney disease protein 1).
DE PKD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95254638; PubMed=7736581;
RA Gluecksmann-Kuis M.A., Tayber O., Woolf E.A., Bougueleret L.,
RA Deng N., Alperin G.D., Iris F., Hawkins F., Munro C., Lakey N.,
RA Duyk G., Schneider M.C., Geng L., Zhang Z., Torosian S.,
RA Reenders S.T., Bork P., Pohlschmidt M., Loehning C., Kraus B.,
RA Nowicka U., Leung A.L.S., Frieschaut A.-M.;
RT "Polycystic kidney disease: the complete structure of the PKD1 gene
and its protein.";
RT Cell 81:289-298(1995).
RN [2]
RP SEQUENCE OF 2769-4303 FROM N.A.
RX MEDLINE=94273192; PubMed=8004675;
RA Ward C.J., Peral B., Hughes J., Thomas S., Gamble V.,
RA Maccarthy A.B., Sloane-Stanley J., Buckle V.J., Kearney L.,
RA Higgs D.R., Ratcliffe P.J., Harris P.C., Roeliffema J.H.,
RA Spruit L.L., Saris J.J., Dauwerse H.G., Peters D.J.M.,
RA Breuning M.H., Nellist M., Brook-Carter P.T., Maheshwar M.M.,
RA Cordeiro I., Santos H., Cabral P., Sampson J.R., Janssen B.,
RA Hesselting-Janssen A.L.W., van den Ouweland A.M.W., Eussen B.,
RA Verhoeve S., Lindhout D., Halley D.J.J.;
RT "The polycystic kidney disease 1 gene encodes a 14 kb transcript and
lies within a duplicated region on chromosome 16. The European
RT Polycystic Kidney Disease Consortium.";
RT Cell 77:881-894(1994).

[3]
 RN STRUCTURE BY NMR OF 275-354.
 RX MEDLINE=99107746; PubMed=9889186;
 RA Bycroft M., Bateman A., Clarke J., Hamill S.J., Sandford R.,
 RA Thomas R.L., Chochia C.;
 RT "The structure of a PKD domain from polycystin-1: implications for
 RT polycystic kidney disease.";
 RL EMBO J. 18:297-305(1999).
 RN [4]
 RP VARIANT ADPKD 3748-ARG--VAL-3752 DEL, AND VARIANT ASP-3632.
 RX MEDLINE=96108969; PubMed=8554072;
 RA Peral B., San Millan J.L., Ong A.C.M., Gamble V., Ward C.J.,
 RA Strong C., Harris P.C.;
 RT "Screening the 3' region of the polycystic kidney disease 1 (PKD1)
 RT gene reveals six novel mutations.";
 RL Am. J. Hum. Genet. 58:86-96(1996).
 RN [5]
 RP VARIANTS ADPKD P-2993; R-3016 AND V-3511, AND VARIANTS M-3510 AND
 RP F-4190.
 RX MEDLINE=97342914; PubMed=9199561;
 RA Peral B., Gamble V., Strong C., Ong A.C.M., Sloane-Stanley J.,
 RA Zarres K., Winears C.G., Harris P.C.;
 RT "Identification of mutations in the duplicated region of the
 RT polycystic kidney disease 1 gene (PKD1) by a novel approach.";
 RL Am. J. Hum. Genet. 60:1399-1410(1997).
 RN [6]
 RP VARIANT ALA-4058.
 RX MEDLINE=97295081; PubMed=9150733;
 RA Constantiniides R., Xenophontos S.L., Neophytou P., Nomura S.,
 RA Pierides A., Constantinou-Deltas C.D.;
 RT "New amino acid polymorphism, Ala/Val4058, in exon 45 of the
 RT polycystic kidney disease 1 gene: evolution of alleles.";
 RL Hum. Genet. 99:644-647(1997).
 RN [7]
 RP VARIANTS T-2760; P-2761; V-2763; T-2764; Q-2791; T-2826; L-3008 AND
 RP L-3064.
 RX MEDLINE=97449169; PubMed=9285784;
 RA Watnick T.J., Piontek K.B., Cordal T.M., Weber H., Gandalph M.A.,
 RA Qian F., Lens X.M., Neumann H.P.H., Germino G.G.;
 RT "An unusual pattern of mutation in the duplicated portion of PKD1 is
 RT revealed by use of a novel strategy for mutation detection.";
 RL Hum. Mol. Genet. 6:1473-1481(1997).
 RN [8]
 RP VARIANT ADPKD THR-3678.
 RX MEDLINE=97401939; PubMed=9259200;
 RA Turco A.E., Rossetti S., Bresin E., Englisch S., Corra S.,
 RA Pignatti P.F.;
 RT "Three novel mutations of the PKD1 gene in Italian families with
 RT autosomal dominant polycystic kidney disease.";
 RL Hum. Mutat. 10:164-167(1997).
 RN [9]
 RP VARIANT ADPKD ASP-4032, AND VARIANT VAL-4045.
 RX MEDLINE=98180892; PubMed=9521593;
 RA Daniells C., Maheshwar M.M., Lazarou L., Davies F., Coles G.,
 RA Ravine D.;
 RT "Novel and recurrent mutations in the PKD1 (polycystic kidney
 RT disease) gene.";
 RL Hum. Genet. 102:216-220(1998).
 RN [10]
 RP VARIANT ADPKD MET-3375.
 RX MEDLINE=99118881; PubMed=9921908;
 RA Koptides M., Constantiniides R., Kyriakides G., Hadjigavriel M.,
 RA Patsalis P.C., Pierides A., Deltas C.C.;
 RT "Loss of heterozygosity in polycystic kidney disease with a missense
 RT mutation in the repeated region of PKD1.";
 RL Hum. Genet. 103:709-717(1998).
 RN [11]
 RP VARIANTS ADPKD L-324 AND S-845, AND VARIANTS R-1399 AND L-1786.
 RX MEDLINE=99294580; PubMed=10364515;
 RA Thomas R.L., McConnell R., Whittaker J., Kirkpatrick P., Bradley J.,
 RA Sandford R.;
 RT "Identification of mutations in the repeated part of the autosomal
 RT dominant polycystic kidney disease type 1 gene, PKD1, by long-range
 PCR.";
 RL Am. J. Hum. Genet. 65:39-49(1999).
 RN [12]
 RP VARIANTS ADPKD P-2392 AND F-2423, AND VARIANTS R-1399; Q-2548 AND
 RP R-2638.
 RX MEDLINE=20046890; PubMed=10577909;
 RA Watnick T., Phakdeekitcharoen B., Johnson A., Gandalph M., Wang M.,
 RA Briefel G., Klinger K.W., Kimberling W., Gabow P., Germino G.G.;
 RT "Mutation detection of PKD1 identifies a novel mutation common to
 RT three families with aneurysms and/or very-early-onset disease.";
 RL Am. J. Hum. Genet. 65:1561-1571(1999).
 RN [13]
 RP VARIANTS ADPKD 3994-L--F-3996 DUPL; G-4136 AND C-4154, AND VARIANTS.
 RX MEDLINE=20441957; PubMed=10987850;
 RA Perrichot R.A., Mercier B., Simon P.M., Whebe B., Clodes J., Ferec C.;
 RT "DSGE screening of PKD1 gene reveals novel mutations in a large cohort
 RT of 146 unrelated patients.";
 RL Hum. Genet. 105:231-239(1999).
 RN [14]
 RP VARIANTS ADPKD 3748-R--V-3752 DEL AND L-4132 DEL, AND VARIANT V-4045.
 RX MEDLINE=20112427; PubMed=10647901;
 RA Afzal A.R., Hand M., Ternes-Pereira E., Saggar-Malik A., Taylor R.,
 RA Jeffery S.;
 RT "Novel mutations in the 3' region of the polycystic kidney disease 1
 RT (PKD1) gene.";
 RL Hum. Genet. 105:648-653(1999).
 RN [15]
 RP VARIANTS ADPKD PRO-4225 AND TRP-4276.
 RX MEDLINE=99217041; PubMed=10200984;
 RA Badenach C., Torra R., San Millan J.L., Lucero L., Mila M.,
 RA Estivill X., Darnell A.;
 RT "Mutation analysis within the 3' region of the PKD1 gene.";
 RL Kidney Int. 55:1225-1233(1999).
 RN [16]
 RP VARIANTS ADPKD MET-2250 AND TRP-2329, AND VARIANTS CYS-2379; LEU-3066;
 RP VAL-3139 AND LEU-3193.
 RX MEDLINE=20311156; PubMed=10854095;
 RA Perrichot R., Mercier B., Quere I., Carre A., Simon P., Whebe B.,
 RA Clodes J., Ferec C.;
 RT "Novel mutations in the duplicated region of PKD1 gene.";
 RL Eur. J. Hum. Genet. 8:353-359(2000).
 RN [17]
 RP VARIANTS ADPKD PRO-2921 AND MET-3375, AND VARIANT LEU-3066.
 RX MEDLINE=20382887; PubMed=10923040;
 RA Koptides M., Mean R., Demetriou K., Constantiniides R., Pierides A.,
 RA Harris P.C., Deltas C.C.;
 RT "Screening of the PKD1 duplicated region reveals multiple single
 RT nucleotide polymorphisms and a de novo mutation in Hellenic
 RT polycystic kidney disease families.";
 RL Hum. Mutat. 16:176-176(2000).
 RN [18]
 RP VARIANTS ADPKD GLN-3719 AND PRO-3852, AND VARIANT VAL-4045.
 RX MEDLINE=20514565; PubMed=11058904;
 RA Agutari G., Savelli S., Garbo M., Bozza A., Augello G., Penolazzi L.,
 RA De Paoli Vitali E., La Torre C., Cappelli G., Piva R., del Senno L.;
 RT "Novel splicing and missense mutations in autosomal dominant
 RT polycystic kidney disease 1 (PKD1) gene: expression of mutated
 RT genes.";
 RL Hum. Mutat. 16:444-445(2000).
 RN [19]
 RP VARIANTS ADPKD SER-1166; GLU-1956; CYS-2408 AND GLY-2442--2443 INS,
 RP AND VARIANTS HIS-1995 AND ASN-2604.
 RX MEDLINE=20467506; PubMed=11012875;
 RA Phakdeekitcharoen B., Watnick T.J., Ahn C., Whang D.-Y., Burkhardt B.,
 RA Germino G.G.;
 RT "Thirteen novel mutations of the replicated region of PKD1 in an Asian
 RT population.";
 RL Kidney Int. 58:1400-1412(2000).
 RN [20]
 RP VARIANTS ADPKD TRP-3753 AND ASN-3815.
 RX MEDLINE=20275386; PubMed=10729710;
 RA Kim U.K., Jin D.K., Ahn C., Shin J.H., Lee K.B., Kim S.H., Chae J.J.,
 RA Hwang D.Y., Lee J.G., Namkoong Y., Lee C.C.;

RT "Novel mutations of the PKD1 gene in Korean patients with autosomal
RT dominant polycystic kidney disease.";
RL Mutat. Res. 432:39-45(2000).
RN [21]
RP VARIANTS ADPKD Q-13; F-75; C-139; 1992-F-T-1993 DELINS L; 2220-R--P-
RP 2224 DEL; D-2336; D-2752; 2762-ILMR-2765 DUPL; M-2768; K-2771; P-2816;
RP S-2858; 3012-T-Y-3017 DEL AND 3748-L-R-3752 DEL, AND VARIANTS S-
RP 2674; M-2708; T-2734; L-2735; C-2765; M-2782; R-2814; G-2888; I-2905;
RP D-2966 AND L-3066.
RX MEDLINE=21063179; PubMed=11115377;
RA Rossetti S., Srmecki L., Gamble V., Burton S., Sneddon V., Peral B.,
RA Roy S., Bakkaloglu A., Komel R., Winealis C.G., Harris P.C.;
RT "Mutation analysis of the entire PKD1 gene: genetic and diagnostic
RT implications.";
RL Am. J. Hum. Genet. 68:46-63(2001).
RN [22]

Query Match 26.9%; Score 208; DB 1; Length 4303;
Best Local Similarity 2.2%; Pred. No. 50;
Matches 87; Conservative 19; Mismatches 41; Indels 3862; Gaps 27;

QY 2 AGG-----ILH----- 7
DB 23 AGGPRCCGCEPPCLCGPAPGAACRVNCSGRGLRTGLPALRIPADATELDVSHLLRAL 82
QY 8 ----- 7
DB 83 DVGLLANLSALAELDISNNKISTLEEGIFANLNFNLSEINLSGNPFBCDCGLAWLPQWAE 142
QY 8 ----- 7
DB 143 QQVVVQPEAATCAGPGSLAGOPLLGPIPLDSCGEEYVACLPNSSGTVAAVFSAAHE 202
QY 8 ----- 7
DB 203 GLLOPEACSAFCSTGGLAALSEQWCLCGAAQSSASFACLSLCSGGPPAPPTCRGP 262
QY 8 ----- 7
DB 263 TLQHVFPASGATLVGPHGLASGQLAAFHIAAPLVTDTRWDFGDSAEVDAAGPAAS 322
QY 8 -----LELL----- 11
DB 323 HRYVLPGRYHTAVLALGAGSALLGTDVQVEAAPALELVCPSSVQSDSLDSIQNRGG 382
QY 12 ----- 11
DB 383 SGLEAAYSIVALGEEPARAVHPLCPDTEIFPGNGHCYRLVVEKAAWLQAEQOCQAWAGA 442
QY 12 -----VAVGP----- 16
DB 443 ALAMVDSPAVQRFVLSRVTRSLDWIGFSTVQGVGPAPQGEAFSLESQNLPGEPHP 502
QY 17 ----- 16
DB 503 ATAHCVRLLGFTGCNCTDLCSAPHYVCELPQGGPVQDAENLLVGAPSGDLQGLTPLAQ 562
QY 17 -DVFOAQHEDTERVV-----LTNLNIGAEILRDP----- 44
DB 563 QDGLSAPHEPVEVWVFPGLRLSREAFLLTAEFTGQLRRPAQLRQYRLLTAGTPENG 622
QY 45 ----- 44
DB 623 SEPESSPDNRQTOLAPACMPGGRWCPGANICLPDASCHPQACANGCTSGPLPGAPYAL 682
QY 45 ----- 44
DB 683 WREFLFSVPAGPPAQYSVTLHGQDVLMLPGDLVQLQHDAGPGALLHCSAPGHPGPRAPY 742
QY 45 ----- 44
DB 743 LSANASSWLPHLPAQLEGTWGCPACALRLLAQREQLTVLLGLRPNPGLRLPGRYEVRAEV 802

QY 45 ----- 44
DB 803 GNGVSRHNILSCSPDVVSPVAGLRIYAPRDCGRLYVPTNGSALVLQVDGSANATATARP 862
QY 45 --SLGAQFR-----VHLVKV----- 58
DB 863 GGSLSARFENVCPALVATEVPACPMETNDTLFSVVALPMLSEGEHVVDVVVENSASRANL 922
QY 59 ----- 58
DB 923 SURVTAEEPICGLRATPSPPEARVLOGVLVRYSPVVEAGSDMVFRMTINDKQSLTFQNVVF 982
QY 59 ----- 58
DB 983 NVIYQSAAVFKLSLTASNHVSNVTYVNMVRMNRMOGLQVSTVPAVLSPNATLALTAG 1042
QY 59 ----- 58
DB 1043 VLVDSAVEAFLWTFDGEQALHQFPYPNESPVPDPSPAQVLVEHNVTHTYAAPGEYL 1102
QY 59 ----- 58
DB 1103 LTVLASAPENLTQQVPVSVRASLPNAVGVSDGVLVAGRPVTFYPHPLSPGGVLYTWD 1162
QY 59 ----- 58
DB 1163 FGDGSPVLTSQSPAANHTYASRGTYHVRLEVNNTVSGAAQADVRVFEELRGLSVDMSLA 1222
QY 59 -----ILTEPE----- 64
DB 1223 VEQGAPVVVSAAVOTGDNIWTFTDMGDGTVLSGPEATVEHVYLRAGNCTVTYVAGSPAGH 1282
QY 65 -----CA 66
DB 1283 LARSHLVLVLELVRVEPAACIPQPDARLTAYTGNPAHYLFDWTFDGSNTTVRG 1342
QY 67 PNITANLTSS-----LLSVC----- 81
DB 1343 PTVTHNFTSRGTFPLALVLSRRNAHVFTSICVEPEVGNVTLQPEROFVLGDEAWLVA 1402
QY 82 ----- 81
DB 1403 CAMPPFPVRYTWFDTBEEAAPTARGPEVTFIYRDPGSLYTVTASNINISAANDSALVEV 1462
QY 82 ----- 81
DB 1463 QEPVLVTSIKVNGSLGLELOQPYLPSAVGRGRPASYLWDLDCGGWLEGPEVTHAYNSTGD 1522
QY 82 -----GM-----SQT----- 87
DB 1523 FTVRVAGNVEVSRSEAWLNVTVKRRVRGLVWNASRTVYPLNGSVSFSTSLGAGSDVRYSM 1582
QY 88 ----- 87
DB 1583 VLCDRCTPIPGGPTISYTFRSVGTENIIVTAENEVGSQADSIFVYVYQLIEGLVVGGR 1642
QY 88 ----- 87
DB 1643 YPPTNHTVLOQAVRDGTVNSYSWTAWDRGRPALAGSKGPSLTVLEAGTYHVQLRATNM 1702
QY 88 ----- 87
DB 1703 LGSAWADCTMDFEVPVGLMWVAASPNPAAVNTSVTSLAELAGSGVYVYVWTSLEEGLSWET 1762
QY 88 ----- 87
DB 1763 SEPFTTHSFPTGLHLVTMTAGNPLGSANATVEVDVQVPVSGLSIRASEPGGSFVAAGSS 1822
QY 88 ----- 87
DB 1823 VPFWQLATGTVNSHCWAVPGSSKRGPHVTMVPDAGTFSIRLNASNAVSHVSATYNTLT 1882
QY 88 -----NPE----- 90


```

Db 1883 AEEPIVGLVWASSKVAPOQLVHFQIILLAGSAVTERLQVGGANPEVLPGPRESHSPFR 1942
Qy 91 ----- 90
Db 1943 VGDHVSVRGKNHVSMAQAQVRIIVLEAVSGLQVNPCEPGIATGTERNFTARVQGRSV 2002
Qy 91 ----- 90
Db 2003 AYANYFSLQKVOGDSLVILSGRDVYTPVAAGLLEIQVAFNALGSENRTLVLEVDQAVQ 2062
Qy 91 -----PG----- 96
Db 2063 YVALQSGPCFTNRSQAFAATSPSPRRVAYHWDGDSGPGQDTDEPRAEHSYLRPGDYRV 2122
Qy 97 -----HADL-----VLYI 104
Db 2123 QVNASNLVSFFVAQATVTVOVLACREPEVDVVLQVLMRRSQRNYLEAHVDLDCVTYQ 2182
Qy 105 TR----- 106
Db 2183 TEYRWEVYRTASCORPCGRPARVALPGVDVSRPRLVPLRLALPVGHYCFVFWVSGDTPLT 2242
Qy 107 -----FDLELPDG----- 114
Db 2243 QSIQANVTVAPERLVPPIIEGGSYRVMSDTRDLVLDGSESYDPNLEDGQTPLSFHWACVA 2302
Qy 115 -----NRQV----- 118
Db 2303 STOREAGGCALNFGPGSSVTIPRERLAAGVEYFTSLTVMKAGRKEEATNOTVLIIRSGR 2362
Qy 119 ----- 118
Db 2363 VPIVSLBCVSKAAQVTEVRSSVYVLEGRCLNCSSGSKGRWAARTFSNKTVLVDDETTT 2422
Qy 119 -----RGVTO-----LGGAC----- 128
Db 2423 STGSAGMLVLRRCVLDGEGYFTTLTVLGRSGEEGCASIRLSPNRPLGSGCRLPLG 2482
Qy 129 ----- 128
Db 2483 AVHALTKVHFECTGMHDAEDAGAPLVYALLRRCRQGHCEFCVYKGSLSYGAVLPPG 2542
Qy 129 ----- 128
Db 2543 FRPHFEVGLAVVQDQLGAAVVALNRSIAITLPEPNSGATGLTVMLHGLTASVLPGLLRQ 2602
Qy 129 ----- 128
Db 2603 ADPOHVIEYSLAVTVLNEYERALDVAAEPKHERQRAQIRKNITETLVSLRVHTVDDIQ 2662
Qy 129 -----SPT----- 131
Db 2663 QIAAALAQCMPSRELVCRSCLKOTLHKLEAMMLILOAETTAGTVTPTAIGDSILNITGD 2722
Qy 132 ----- 131
Db 2723 LIHLASSDVRAPQPSSELGAESPMSRMVASQAYNLTSALMILMRSRVLNESEPLTLAGEIV 2782
Qy 132 ----- 131
Db 2783 AQGRSDPRSLCYGAPGPGCHFSIPEAFSGALANLSDVVQLIFLVDSNPPFPFGYISNY 2842
Qy 132 ----- 131
Db 2843 TVSTKVASMAFQTOGAQIPIERLASERAITVKVPNNSDWAARGHRSSANSVYVQPQ 2902
Qy 132 ----- 131
Db 2903 ASGVAVTLSSNPAAGLHLQLNYTLDDGHYLSEEPYLAVYLHSEPRNEHNCASRR 2962
Qy 132 -----WS----- 133

```

RESULT 15
FATH_HUMAN

```

Db 2963 IRPESLQAGADHRPYTFTFISPSRDPAGSYHNLSSHFRMSALQVSVGLYTSLCQYFSEED 3022
Qy 134 ----- 133
Db 3023 MVRTEGELLPLEETSPROAVCLTRHLTAFCASLFPVPSHVRFVPEPTADVNYIVMLTCA 3082
Qy 134 -CLIT----- 137
Db 3083 VCLVTYVMMAAILHLKLDQDASRGRAIPFCQGRPKYIILVKTGMRGSGOTTAAHVIGML 3142
Qy 138 ----- 137
Db 3143 YGVDSRSGHRHLDCDRAFHRLSLDIIFRIATPHSLGSLYMKIRVMHDKGLSPAWFLOHVIV 3202
Qy 138 ----- 137
Db 3203 RDLQARSAPFLVNDMLSVETEANGGLVEKEVLAASDAALLRFRLLVAELQRFQDKHI 3262
Qy 138 ----- 137
Db 3263 WLSIWDPRPSRFRTRIORATCCVLLICLPLGANAVMYGAVGDSAYSSTGHVSRLSPLSVD 3322
Qy 138 ----- 137
Db 3323 VAVGLVSSVVVYPVYLAILFLFRMSRSKVAGSPSPAGQQVLDIDSLDSSVLDS8FLT 3382
Qy 138 ----- 137
Db 3383 FSGLHAEQAFVGOMKSDFLDODSKSLVWPCSGEGTISWPDLSDPSIVGSLNRLQARGQA 3442
Qy 138 ----- 137
Db 3443 GHGLGPEEDGFSLASPYSPAKSFSASDEDLIQQVLAEGVSSPAPTQDTHMETDLLSSLS 3502
Qy 138 ----- 137
Db 3503 TPGEKTELTALQRLGELGPPSPGLNWEQPOARLSRTGLVEGLRKLPAWCAASHGLS 3562
Qy 138 ----- 137
Db 3563 LLLVAVAVAGVWGAGSPFPVGVAMVLLSSASFLASFLGWEPKVLLEALYFSLVAKRL 3622
Qy 138 ---ED----- 139
Db 3623 HPDEDDTLVESPAVTPVSAARVPRVPPHGFALFLAKEEARVKVRLHGMRLSLVYMLFL 3682
Qy 140 ----- 139
Db 3683 VTLASVGDASCHGHAYRLOSALIKOELHSRAFLAITRSEELWPMAHVLLPYVHGNQSSP 3742
Qy 140 ----- 139
Db 3743 ELGPPRLQVRLOEALYPDPGPRVHTCSAAGFSSTDYDVGWESPHNGSGTWAYSAPDL 3802
Qy 140 ----- 139
Db 3803 LGAWSGCAVYDSGGYVOELGSLSESRDLRFLQHLNHNDRSRAVFLLETRYSPAVG 3862
Qy 140 ----- 139
Db 3863 LHAATVLRLEFFPAAGRALAALSVRPFALRRLSAGLSPLLTSLVCLLLFAVHFAVAEARTW 3922
Qy 140 -----TCPD----- 143
Db 3923 HREGRWVRLRGAWARWLLVALTAATLVRLAOLGAADROWTRFVGRPRRFTSFDOVAH 3982
Qy 144 -----LGYT1 148
Db 3983 VSSAARGLAASLLFLLLVKAQHVRVQWVSFGKTLCLRALPELLGVTL 4031

```


Qy	7	-----HLEL-----	10	Qy	35	-----	34
Db	416	NTGLISILEPVKQQAHFLEVTTSDRKASTKVLVKVLGANSNPPEFTQAYKAAFDEN	475	Db	1496	SRDPLSLKFLRDPATGSLYTSBKLOHEAVSPAHLTMVRDQDQVPVKRNFARIVVNSDT	1555
Qy	11	-----LVAGPD-----	17	Qy	35	-----NIGABELLR	42
Db	476	VPIGTTIMSLSAVDPEDEGENGYVTVSYIANLHVPPAIDHFTCAVSTSENLDYELMPRYVT	535	Db	1556	NDIAPHTASSYKGRVYESAAVGSVVLQVLTALDKDKNAEVLYSIESGIGNIGNSFMI	1615
Qy	18	-----	17	Qy	43	DPSLGA-----	48
Db	536	LRIRASDWGLPYRREVEVLATITLNNLNDNTPLFEKINCEGTIPRDLGVGEQIITVSAID	595	Db	1616	DPVLGSIKTAKELDORSQAQAEYDLWKATDKGSPMSEITSVRIFVTIADNASPKFTSKEY	1675
Qy	18	-----	17	Qy	49	-----	48
Db	596	ADELQVQYQIEAGNELDLFSLNPNSGVLSLKRSIMDGLGAKVSHFSLRITATDGENPAT	655	Db	1676	SVELSETVSIGSFVGMVTAHSQSSVVVEIKDGTGDAFDINPHSGTIIITQKALDFETLPI	1735
Qy	18	-----	17	Qy	49	-----QPRVHL-----	61
Db	656	PLYINITVAASHKLVLQCEBETGVAKMLAEKLLQANKLHQGEVEDIFFDSHSYNAHIPQ	715	Db	1736	YTLIIOGTNMAGLSTNTVTLVHLQDENNDNAPVFMQABEYTGGLISESASINSVVLTDNRNVL	1795
Qy	18	-----	17	Qy	62	-----EP-----	63
Db	716	FRSLPTGIVKENQPVGSSVIFMNSTDLDTGFGNKLVIYVSGGNEDSCFMIDMETGMLK	775	Db	1796	VIRAADADKDSNALLVYHIVEPSTHTYPAIDSSGTGAIHTVLSLDYBETSIFHFTVOVHDM	1855
Qy	18	-----	17	Qy	64	-----EGAPNIT-----	70
Db	776	ILSPLDRETTDKYTLNITVYDLGIPQAAWRLLHVVVVVDANDNPPFLOESYFVEVSEDK	835	Db	1856	GTPLRAEYAAANTVHVHIDINDCPPVFAKPLYEASLLLPYKGVKVITVNATDADSAFS	1915
Qy	18	-----	17	Qy	71	-----ANLTS-----	75
Db	836	EVHSEIIQVEATDKDLGNHGVTSILTDTFSDSVTVGVVNIARPLDRELQHEHSLKI	895	Db	1916	QLIYSITEGNEGKFSMDYKGTALTQVNTTQLRSRVELTVRASDGRFAGLTSVKINVKES	1975
Qy	18	-----	17	Qy	76	-----	75
Db	896	EARDOAREEPOLFSTVVVVKVSLDVDNDNPPTFIPNVYRKVREDLPECTVIMMLEAHDPP	955	Db	1976	KESHLKFTQDVYSAVVKENSTEATLAVITAIGSPINEPLFYHILNPDPRFKLSRTSGVL	2035
Qy	18	-----	17	Qy	76	-----	75
Db	956	LGQSGOVYSLLDHGEGNFVDYKLSGAVRIVQQLDFEKKQVYNLTVRAKDKGKPVLSST	1015	Db	2036	STGTPTPDREQQBAFDVWVEVIEBKHPSAVAHVVVVKVIVEDQNDNAPVFNLPYAVVKV	2095
Qy	18	-----VF-----QAHOED-----	25	Qy	76	-----	75
Db	1016	CYVEVEVDNENLHPVPFSSFVEKGTVKEDAPVGSVMTVSAHDEADAGRGEIRYSIRD	1075	Db	2096	DTEVGHVIRYVTAVDGRNGRGEVHYLKEHHEHFOIGLGEISLKKQFELDTLNKEYLV	2155
Qy	26	-----TERYVLT-----	32	Qy	76	-----	75
Db	1076	GSVGVPKIGBETGVIETSDRLDRESTSHYLTVFATDQGVVPLSSFIEIYIEVEDVNDN	1135	Db	2156	TVVAKDGGNPAPSAEVIPIVTVNKNKAMPVEKPFYSAETAESIOVHSPPVHVQANSPEGL	2215
Qy	33	-----	32	Qy	76	-----	75
Db	1136	APQTSPPVYPEIMENSPKDVSVVQIEAFDPDSSNDKLMYKITSGNPQGFPSHPKTGL	1195	Db	2216	KVFYSITDGDPPFSQFTINFNTGVINVIAPLDPEAHPAYKLSIRATDLSLTCAHAEVFVDII	2275
Qy	33	-----	32	Qy	76	-----	75
Db	1196	ITTSRKLDRQODEHILEVTVDNGSPPKSTIARVIVKILDENDNKPQFLQKFKIRLP	1255	Db	2276	VDDINDNPPVFAOQSYAVTLSEASVIGTSVVQVRATDSDBEPNRGISYOMFGNHSKSHDH	2335
Qy	33	-----	32	Qy	76	-----	75
Db	1256	EREKPDRENNARREPLYRIATDKDEGNABEISYIEDGNEHGKFFIEPKTVGVVSKRFS	1315	Db	2336	FHVDSTGLISLRLTLDYEOSROHTTFVRAVDGGMPTLSSDVIIVTDVDTLNGNPPLPEQ	2395
Qy	33	-----	32	Qy	76	-----	75
Db	1316	AAGEYDILSIKAVONGRPQKSTTRLHIEWISKPKQSLPISFEESFTFTVMESDPVAH	1375	Db	2396	QIYEARISEHAPHGHFVTCVKAYDADSSDIDKLOYLSILSGNDHKHFVIDSATGIITLSNL	2455
Qy	33	-----NL-----	34	Qy	76	-----	75
Db	1376	MIGVISVEPPGIPLWFDITGNGYDSHPDVKDGTGTIIIVAKPLDAEQKSNYLTVEATDGT	1435	Db	2456	HRHALKPPYSLNLSVSDGVPRSSSTQVHVTVIGGNLHSPAFLONEYEVELAENAPLHTLVM	2515
Qy	35	-----	34	Qy	76	-----	75
Db	1436	TTILTQVFIKVIDNDRHROPFSTSKYEVVIPEDTAPETELIQISAVDQDEKGNKLIYTLOS	1495	Db	2516	EVKTTDGDGSGIYGHVTHIYVNDPAKDRFYINERGOIFLTLEKLDRETPAEKVISVRLMAKD	2575
				Qy	76	-----	75

Db 2576 AGGKVAFTVNVILTDNDNAPQFRATKYEVNIGSAAKGTSWKASDADEGSNADITY 2635
QY 76 ----- 75
Db 2636 AIEADSESVKENLKNLSGVITTKESLIGLENEFFTFVRAVDNGSPSKESVVLVYVKI 2695
QY 76 ----- 75
Db 2696 LPPEMQLPKFSEBFFYFTVSEDPVPGTEIDLIRAEHSGTVLYSLVKNGTNPESNRDESFI 2755
QY 76 ----- 81
Db 2756 DRQSGRLKLEKSLDHETTKWYQFSILARCTQDDHEMVASVDVSIQVKDANDNSPVFESSP 2815
QY 82 ----- 83
Db 2816 YEAFIVENLPGSRVQIRASDADSGTNGQVMYSLDOSQOSVEVIESFAINMETGMITTLK 2875
QY 84 ----- 83
Db 2876 ELDHEKRDNYQIKVASDHGEKIQLSSTAIVDVTVTDVNDSPRFTAEIYKGTVSEDDPO 2935
QY 84 ----- 88
Db 2936 GGVIALSTTDADSEINRQVTFITGDPGLQFAVETIQNEWKVYVKKPLDREKRDNYL 2995
QY 89 ----- 91
Db 2996 LTITADGTFFSKAIVEVKVLDANDNSPVCEKTLYSDDTIPEDVPGKLIMQISATDADIR 3055
QY 92 ----- 91
Db 3056 SNAEITYTLGSAEKFKLNPDTGELKSTPLDREBQAVYHLLVRATDGGRCQASIVV 3115
QY 92 ----- 96
Db 3116 TLEDVNDNAPEADPYAITVENTEPGTLTRVQATDADAGLNKILYSLIDSADGQPS 3175
QY 97 ----- 96
Db 3176 INELSGIIQLEKPLDRELOAVYTLKAVDQGLPRRLTATGTVIVSVLDINDNPPVFEYR 3235
QY 97 ----- 96
Db 3236 EYGATVSEDILVCTEVLQVVAASRDIEANAEITYSIISGNECHKFSIDSKTGAFFIENL 3295
QY 97 ----- 96
Db 3296 DYESSHEYVLTVEATDGGTSLSDVATVNVNVTINDNTPVFSQDTYTTVISDAVLEQS 3355
QY 97 ----- 96
Db 3356 VITVMADADAGPSNSHIHYSIIDNQGOSFTIDPVRGEVKVTKLLDRETISGYTLTVQAS 3415
QY 97 ----- 96
Db 3416 DNGSPRVNTTVDVSDVNDNAPVFSRGNYSVIIQENKPVGFSVLQVLVTTDEDSHNG 3475
QY 97 ----- 96
Db 3476 PPEFFTVITGNDEKAFEVNPQGVLLTSSAIKRKEKDHLYLLQVKVADNGKPOLSSLTYIDI 3535
QY 97 ----- 96
Db 3536 RVIEESIYPAILPLEIFITSSGEYSGVIGIKIHATDQDVYDTLTYSLDPOMDNLFVS 3595
QY 97 ----- 96
Db 3596 STGCKLIAHKKLDIGVLLNVSVTDGKFTTVADITVHIROVTOEMLNHTIAIRFANLTPE 3655
QY 97 ----- 105

Db 3656 EFVGDYWRNFORALRNILGVRRNDIQIVSLOSSEPHPLDVLVLFVEKPSAQISTKOLLH 3715
QY 106 ----- 105
Db 3716 KINSSVTDIEEIIIGVRILNVFQKLCAGLDCPWKFCDEKVSVDSEVMSTHSTARLSFVTPR 3775
QY 106 ----- 105
Db 3776 HRAAVCLCKEGRCPVHHGCCDDPCPEGSECVSDPWEKHKTCVPCSGRFOCGSSSMT 3835
QY 106 ----- 109
Db 3836 LTGNSYVKYRLTENENKLEMLTMRRTYTHAVVMYARGTDYSILEIHHGRLOVKFDCG 3895
QY 110 ----- 109
Db 3896 SGPGIVSQSTQVNDGOMHVALEVNGNYARLVLDQVHTASGTAPGTLKTLNLDNYVFFG 3955
QY 110 ----- 116
Db 3956 GHIROQTRHGRSPQVNGFRGCMDSIYLNGOELPLNSKPRSYAHIEESVDVSPCCFLTA 4015
QY 117 ----- 116
Db 4016 TEDCASNPONGGVCNPSAGGYCKSALYIGTHCEISVNPCSSNPCLYGGTCVVDNGG 4075
QY 117 ----- 128
Db 4076 FVOCRGLYTQRCQLSPYCKDEPKNGGTCFDSLDGAVCOCDSGFRGRCQSDIDECG 4135
QY 129 ----- 128
Db 4136 NPCLHGALCENTHSGYHCNCSHEYRGHCEADAPQYVSTPMNIGLAEGIGIVFVAGIF 4195
QY 129 ----- 128
Db 4196 LLVVVFLCRKWI SRKKKHQAEKDXGLGPATAFLORPYFDSKLNKNIYSDIPPOVVRP 4255
QY 129 ----- 128
Db 4256 ISYTPSIPSDSRNLDNRNSFEGSAIPEHPEFTFNPESVHGHRKAVCVSAPNLP PPP 4315
QY 129 ----- 132
Db 4316 SNSPSDSISIQKPSWDFDYDTKVVDLDPCLSKKPLEEKSPQYSARESLSVQSLSFQS 4375
QY 133 ----- 144
Db 4376 ESCDDNGYHMDTSDMMPVPLDIOEPFNYEVIDEOTPLYSADPNAIDTDYYPGGYDIES 4435
QY 145 ----- 144
Db 4436 DEFPPEPDPAADELPLLPPEFSNQFESIHPPRDMPAAGSLGSSSRNRQRPNLQVLPNF 4495
QY 145 ----- 144
Db 4496 YPLDMSEPTKTGTGENSTCREPHAPYPPQYQHRHEAPAVESMPMSVYASTASCSDVSACC 4555
QY 145 ----- 148
Db 4556 EVESEVMMSDYESGDDGHFEVTI 4579

Search completed: March 20, 2003, 12:39:21
Job time : 75 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:36:42 ; Search time 15 Seconds
(without alignments)
290.306 Million cell updates/sec

Title: SEQID_1_15FUSED

Perfect score: 773
Sequence: 1 AAGGILHLELLVAVGPDVFQ.....SPTWSCLITDGTGDLGVTI 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	264	34.2	15281	2	US-08-471-119A-2
2	237	30.7	997	4	US-09-369-364A-7
3	235	30.4	11877	4	US-09-105-537-6
4	216	27.9	5588	4	US-09-036-987A-6
5	216	27.9	5588	4	US-09-370-700-6
6	209	27.0	7257	3	US-09-335-409-5
7	209	27.0	7257	4	US-09-568-102-5
8	209	27.0	7257	4	US-09-567-969-5
9	209	27.0	7257	4	US-09-568-480-5
10	209	27.0	7257	4	US-09-568-486-5
11	209	27.0	7257	4	US-09-568-472-5
12	209	27.0	7257	4	US-09-567-899-5
13	208	26.9	4303	2	US-08-460-751-2
14	207	26.8	4551	3	US-09-320-878-1
15	207	26.8	4613	4	US-09-105-537-31
16	206	26.6	4302	3	US-08-658-136-5
17	206	26.6	4302	4	US-09-052-469-8
18	206	26.6	4339	4	US-09-052-469-6
19	206	26.6	5405	4	US-08-718-388-9
20	204	26.4	1882	4	US-09-369-364A-13
21	204	26.4	10182	4	US-09-134-001C-3159
22	202	26.1	6095	4	US-09-144-085-2
23	200	25.9	4928	4	US-09-036-987A-5
24	200	25.9	4928	4	US-09-370-700-5
25	199	25.7	3898	4	US-08-750-717-2
26	197	25.5	859	4	US-09-369-364A-5
27	197	25.5	5215	4	US-09-105-537-2

28	196	25.4	3782	4	US-09-105-537-4	Sequence 4, Appli
29	196	25.4	4654	4	US-08-476-515A-84	Sequence 84, Appli
30	196	25.4	4655	4	US-08-652-877-84	Sequence 84, Appli
31	196	25.4	4655	4	US-08-652-877-86	Sequence 86, Appli
32	196	25.4	4655	4	US-08-652-877-88	Sequence 88, Appli
33	196	25.4	4655	4	US-08-652-877-90	Sequence 90, Appli
34	195	25.2	3665	2	US-08-222-617A-13	Sequence 13, Appli
35	195	25.2	3712	2	US-08-222-617A-4	Sequence 4, Appli
36	195	25.2	3712	2	US-08-222-617A-25	Sequence 25, Appli
37	195	25.2	3898	2	US-08-876-991-2	Sequence 2, Appli
38	195	25.2	3898	2	US-09-059-853-2	Sequence 2, Appli
39	194	25.1	837	4	US-09-122-126B-2	Sequence 2, Appli
40	193	25.0	967	4	US-09-130-491-2	Sequence 2, Appli
41	193	25.0	4472	2	US-08-804-227C-2	Sequence 2, Appli
42	191	24.7	3072	4	US-09-413-814-93	Sequence 93, Appli
43	191	24.7	3079	4	US-09-413-814-80	Sequence 80, Appli
44	190	24.6	2813	3	US-08-896-449A-2	Sequence 2, Appli
45	190	24.6	2813	3	US-09-132-652-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-471-119A-2
; Sequence 2, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoergerdorfer, Kurt
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07936
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kassenoff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15281 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tolypocladium niveum
; STRAIN: ATCC 34921
; US-08-471-119A-2

Query Match 34.2%; Score 264; DB 2; Length 15281;
Best Local Similarity 0.8%; Pred. No. 0.013;

Matches 102; Conservative 24; Mismatches 21; Indels 12475; Gaps 32;

Qy	1	AAGGIL-----	Qy	19	-----	18
		:				
Db	2098	AAGHLEIGTGGMVLFNLQAGLSYIGLEPSQSAVFVNKAAQTFFGLEGKAQVHGT 2157	Db	3178	PIEGMAEETIRITEILADAKTDDINGLAASOPTAASLAYVFTSGSTGRPKGVMEHR 3237	
Qy	7	-----	Qy	19	-----	18
		6				
Db	2158	AMDTGRLSALSPDLIVNSVAQYFSPSREYLAEVVEALVRIPGVRRIFFGDMRTYATHKDF 2217	Db	3238	GIVRLTKQNTSKLPESFMAHISNLAFDASVWEVFTLLNGGTLVCIDYFTLLESTAL 3297	
Qy	7	-----	Qy	19	-----	18
		-----HLELL----- 11				
Db	2218	LVARVHTNGSKVTRSKVOEVARLEEEELLVDPAPFTSLKESLSEIEHVEILPKNM 2277	Db	3298	EKVFQORVNVALLPALLKQCLDNSPALVKLTSLVLYIGGDRLDASDAKARGLVQTOAF 3357	
Qy	12	-----	Qy	19	-----	18
		11				
Db	2278	KVNNELSSRYGAVLHNRHNQNSRSHKINAESWIDFASSQMDROGLARLLKENKDAE 2337	Db	3358	NAYGPTENTVMSTIYPIAEDPPFINGVPIGHAVSNSGAFYMDQNOQITPPGANGELIVTGD 3417	
Qy	12	-----	Qy	19	-----	18
		11				
Db	2338	STAVFNI PYSKTIVERHTAKSLADHDGDDTHSSIDGVAMWISAAREKASQCPSLDVHDLV 2397	Db	3418	GLARGYTTSSLNTGRFINVDIDGEOVRAYRTGDRVRYRPKLOIEFFGRIDHOVKIRGHR 3477	
Qy	12	-----	Qy	19	-----	18
		11				
Db	2398	OLAEDAGPRVEVSWARQSONGALDVFFHHFQPTENESRALVDPTDYKGOQARS LNRP 2457	Db	3478	IEPAEVEYALLSHDLVTDAAVVVTHSOENQDLEMVGFAARVADVREDESSNOVOEMQTHF 3537	
Qy	12	-----	Qy	19	-----	18
		11				
Db	2458	LORVESRRIEAQVRELOVLLPAYMIPARIVVLQNPLNTSGKVDKELTLRAKVAART 2517	Db	3538	DSIAYADITTDQOGLGRDFMSWTSGSLIKKSQMOEWLDDTMRSLLDSQPPGHVLEV 3597	
Qy	12	-----	Qy	19	-----	18
		11				
Db	2518	PSELVAPRDSIEALICEKFDVLGVEGIDTNFFNVGGHSLATKLAARLSQLNAQIA 2577	Db	3598	GTGTGMVLFNLREGGLOSYYVLEBPSPSATAFVNKAASPPGLEDRIRVEVGTATDIDL 3657	
Qy	12	-----	Qy	19	-----	18
		11				
Db	2578	VKIDFRPVIADLAATIQDTEHNPIPTSYTGPVEQSPAQGLMFLDQLNVGATWYLM 2637	Db	3658	GDDLHAGLVVNSVAQYFSPQDYLAQLVRDLTKVPGVERIFFGDMRSHAINRDFLARAV 3717	
Qy	12	-----	Qy	19	-----	18
		11				
Db	2638	PFVRLRGLVVSALAAALLALEERHETLRTTFIEQEGIGMVIHPFAPKELRVIDVSGE 2697	Db	3718	HALGDKATKAEIQREVVVRMESEDELLVDPAFTSLTTOVENIKHVEILPKRMATNELS 3777	
Qy	12	-----	Qy	19	-----	18
		11				
Db	2698	EESTIQKILEQOTTPPNLASEPGFRLALKTGDEHILSTVMHHAISDGSVDI IQOEI 2757	Db	3778	SYRYAHLVNDLAKPAHKVSPGAWVDAATKMDRDALIRLLRGTKISDHAIATIPNSK 3837	
Qy	12	-----	Qy	19	-----	18
		11				
Db	2758	QFYSAILRGHDFLAQIAPLSIQYRDFATWQRQIFQVAEHRRLAYWTQKADNKP AELL 2817	Db	3838	TIVERTICESYVDLGGDAKOSNDRVSWLSAARSNAVKVASLIDLVDIAQEAQFRVEIS 3897	
Qy	12	-----	Qy	19	-----	18
		11				
Db	2818	TDFKRPMLSGRAEIPVVVDGLIYEKLQDFCRIRQVTAFTVLLAAFRAAHYRMTGTEDA 2877	Db	3898	CARQWSONGALDAVFHHLGPSPOSSHVLIDFLTDHOGRPBEALTNHPLHRAQSRVERQI 3957	
Qy	12	-----	Qy	19	-----	18
		11				
Db	2878	TICTPIANRNRPELEGLIGFFVNTQCMRITVDVEDSFETLVHQVRETTLAAHANQDVPE 2937	Db	3958	RERLOTLLPAYMIPAQIMVLDKPLNANGKVRKOLTQAOQVPAKQVSAVPAPRTEIE 4017	
Qy	12	-----	Qy	19	-----	18
		11				
Db	2938	QIVSNILPGSSDTSRNPVLQMLFALHSQONLQKVRLEGIEEII SIAETTRFDIEFHLYQ 2997	Db	4018	RVLCOEFSVDLVGVDIGIMENFPDLGCHSLMATKLAARISRRLETHSVSVKEIFDHPRVCDL 4077	
Qy	12	-----	Qy	19	-----	18
		11				
Db	2998	EAERLNGSIYAADLVFPETIQSVITIFOGILQKGLGEPDMPVASKALDGGLESURSTGL 3057	Db	4078	VLIVQOGSAPHDPIVSTKYTGPVPOSFAOGRLWFLDQLNFGATWYLMPLAVRLRGAMNVH 4137	
Qy	12	-----	Qy	19	-----	18
		18				
Db	3058	LHPQOTDPCDASVVOIFKQOVAVNPDVIAVRDESTRLSYADLRKSDQVACWLSRRGIA 3117	Db	4138	ALTAALLALERRHELLRTTFYEONGVMQKVNPNVVTETLRIIDLNSGDGDLPTLKKEQT 4197	
Qy	19	-----	Qy	19	-----	18
		18				
Db	3118	PETFAVILAPRSCETIVAILGLVKANLAYLPDVNVNVPASRLAEILSEVSGMLVLVGAET 3177	Db	4198	APFHLETPGRVALLRLPGDYILSVVMHHIISDGSWSDVLVFLQELGQFYSTAVKGDHPL 4257	

```
QY 19 ----- 18
Db 4258 SQTPLPIHYRDFALWQKPTQSEHERQLQYVWEQLVDSAPAELLTDLPRPSILSQAG 4317
QY 19 ----- 18
Db 4318 EMSVTIEGALYKNLEEFRCVRHVRTSFVLLAALRAAAHYRLTGSEDATIGTPIANRRNP 4377
QY 19 ----- 18
Db 4378 EQIIGFVNTQCIRITVNEDETRESLVQVRSATAAFAHQDVPFKEIVSTLLPGSRDAS 4437
QY 19 ----- 18
Db 4438 RNPLVQLMFAVHSQKNLGELKLENAHSEVVPTETITRFDLEPHLFOQDDKLEGSILYSTD 4497
QY 19 ----- 18
Db 4498 LFEAVSQSLLSVFQILRRGLNGPDVPITPLQDGIVDLQORQLLDVQKTEYPRDSSV 4557
QY 19 ----- 18
Db 4558 VDFHEQVSINPDSIALIHSEKLSYAQLDRESDRVRLHRFSFSDTLIAVLAPRSC 4617
QY 19 ----- 18
Db 4618 TIIAFLGILKANLAYPLDVKAPAAIDAIVSSLPGNKLILIGANVTTPKLOEAAIDFVP 4677
QY 19 ----- 18
Db 4678 IRDTFTLTLDGTLODPTIERSAQSALAYAMFTSGTRPKVMVQHRNIVRLVKNVNV 4737
QY 19 -----FOAQ----- 23
Db 4738 AKOPAAARIAHISLAFDASSMEIYAPLLNGGAIVCADYFTIDPQALQETTFQHEIRGA 4797
QY 24 ----- 23
Db 4798 MLPPSLLKQCLVQAPDMISRLDILFAAGDRFSSVDALQORLVGSGVFENAYGPTNTILS 4857
QY 24 ----- 23
Db 4858 TIYNVAENDSFVNGVPIGSAVNSGAYIMDKNQQLVPAGVMGELVVTDGLRGYMDPKL 4917
QY 24 ----- 23
Db 4918 DADRFIQLTVNGSEQVRAVRTGDRVRYRPKDFQIEFFGRMDQQIKIRGHRIEPAEVEQAF 4977
QY 24 ----- 23
Db 4978 LNDGFVEDVAIVIRTPENQEPBMVAVTAKGDNAREEEATTOIEGWEAHFEGGAYANIE 5037
QY 24 -----EDT----- 26
Db 5038 EIESEALGYDFMGWTSMDYDGTIEDKDDEMRLNDTMRSLLDGKPAGRVLEVTGTGIMF 5097
QY 27 -----ERYV----- 30
Db 5098 NLGRSQGLERYIGLEPAPSAAEFVNNAKSPFGLAGRAEVHVGTAADVGTLOGLTSDMAV 5157
QY 31 ----- 30
Db 5158 INSVAQYFPTPEYLAETIKSLVQVPGMKRIYLGDMRSMWNRDFAAARAAYSADNASKD 5217
QY 31 ----- 30
Db 5218 RVROMMELEEBEELLVDPAFTALASQLQDRHQHVEILPKMKATNELSSRYAAVLH 5277
QY 31 ----- 30
Db 5278 ISDEPLIYKIDPEWINFEGSRLTREALAQVLKENENAESVAISNIPYSKTVERHIVR 5337
QY 31 ----- 30

Db 5338 SLDOEDANAPESMDGSDWISAVRTRAQOCHTLSASDLFDAEDAGFRVEVSWARHSQH 5397
QY 31 -----LTN----- 33
Db 5398 GALDAVFHHLKATEDSRVLTKFPTDHQGRPLKSLTNQPLLPQASRRRAELLIREGLQTL 5457
QY 34 ----- 33
Db 5458 PPYMIPSOITLIDRMPLNANGKVDRELARRAKITOKSKPVEDIVPPRNSVEATVCKGFT 5517
QY 34 ----- 33
Db 5518 DVLGVEGITDNFPNLGCHSLMATKLAARLGRQLNTRISVRDVPDQVADLAAVIORNS 5577
QY 34 -----LNIGA----- 38
Db 5578 APHEPIKPADYTGVPVPOSFAQGRLLFQDLQNVGATWYMLPGIRLHCSLRVDALATAISA 5637
QY 39 ----- 38
Db 5638 LEORHEPLRTTFHEEDGVGVVQVQDHRPKDLRIIDLSTOPKDAYLAVLKHEOTTLFOLAT 5697
QY 39 ----- 38
Db 5698 EPGMRVALIRLGBEEHLSIVMHHIISDGSVEVLFDHMRFFYSSALROODPMEOILPLP 5757
QY 39 ----- 38
Db 5758 IOYRDFAAWQKTEQVAEHQRQDYWTEHLADSTPAELLTDLPRPSILSGRANELPLTIE 5817
QY 39 ----- 38
Db 5818 GRLLDKLRAFRCRVHQATPFVILLALRAAAHYRLTGAEDATLGTPTIANRRPELENMIGFF 5877
QY 39 ----- 38
Db 5878 VNTQCMRIAIEENDNPESLVRVRSTATSAFANQDVPFESIVSSLLPGSRDASRNPLQV 5937
QY 39 ----- 38
Db 5938 ILAVHSQDLGKLTLEGLRDEAVD9A18TRPDVEPHLFEHADRLSGSVLYAKELFKLRTI 5997
QY 39 ----- 38
Db 5998 ESVWSVLETLRRLDQPLTPLAVLPLTDGVEIASKGLLDVPRDTPRDANIVEVFOOH 6057
QY 39 ----- 38
Db 6058 VRATPDIAVKDATSILTVAQLDQOSDRLAIWLSRRHMPETLVGLVLAAPRSCETIIAMFG 6117
QY 39 ----- 38
Db 6118 IMKANLAYPLDINSAPAARLSILSAVDGNKVLVLSGVTAPQENPEVEAVGIOELIAG 6177
QY 39 ----- 38
Db 6178 TGLDKTQGSNARPSATSLAYVIFTSGTGPKGVWVEHRSVTRLAKPSNVISKLPQGARV 6237
QY 39 ----- 38
Db 6238 AHLANIAFPDASIWEIATTLNGATLVCLDYHTVLDCTRLKEVPBRESITVVTLMPALLKQ 6297
QY 39 ----- 38
Db 6298 CVAEIPETLAHLDLLTYGGDRVGGHDAMRARSUKVIGMFCGYGPTENTVISTIVEYDADE 6357
QY 39 ----- 38
Db 6358 MFVNGVPIGKTVSNGAYVMDRNOQLVPSGVVGLVVTGDLARGYTDPSLNKURFIYIT 6417
QY 39 ----- 38
```

```
Db 6418 VGESIRAYRTGDRVRPHDLQIEFFGRMDQVKIRGHRIBEPGEVESALLSHNSVQDAA 6477
Qy 39 ----- 38
Db 6478 VVICAPADQDSGAEMVAFVAARNTEDETOEEBAVQVQGWETHFETAAYSEVKDIRQSE 6537
Qy 39 ----- 38
Db 6538 VGNDFMGWTSMYDSEIDKTMHEWLNDRMILDAAREPHVLEIGTGTGMVFNLAACP 6597
Qy 39 ----- 38
Db 6598 GLOCYVGFEPSKAAQFVNDAAOSFPALKDGRSIVHVGTAIDINKAGPIQPRLVVNSVA 6657
Qy 39 ----- 38
Db 6658 QYFPTPEYLFRRVVEALVQIPSVVERIVFGDMRTNAINRDFVSAALHTLGEKANKELVRQM 6717
Qy 39 ----- 45
Db 6718 IVELEANEBELLTDPAFFTSRLRGEKIKHVEILPKTMKATNELSKRYAAVLHVRGSR 6777
Qy 46 ----- 45
Db 6778 EQSTIHQSPNAWIDFAADGLDROTILNLKHKDAGTVAIGNIPYSKTIVERFVNKSL 6837
Qy 46 ----- 45
Db 6838 EDDMEGONSLDGSAAVAAVMAAQCPSLDAMDVKEIAQAGYQVEVSNAROWSONGAL 6897
Qy 46 ----- 49
Db 6898 DAIFHHFPPKEGARTLIEFTDYEGRNVTLTNRPLNSIQSRLGTQIREKQLTLLPPY 6957
Qy 50 ----- 49
Db 6958 MIPSRIWLDQMPVNNCKIDRKELVRRAIVAPKPSAATRAPRNEIEAILRDEFEDVL 7017
Qy 50 ----- 49
Db 7018 GTEVSLDNFFDLGGHSLMATKLAARVSRRLDAHSIKVDQDPVLADLAASIQRESAPH 7077
Qy 50 ----- 49
Db 7078 EPIQRPYTGPAEQSFAQGLWFLDQNLGATWYLMPLAIRIGQLRVAALSAAALFALER 7137
Qy 50 ----- 49
Db 7138 RHETLRTTFBESDGVQIVGEARNSDLRVHDVSTGDDGEYLEVLRREQTVPFDLSSEPG 7197
Qy 50 FRVHLVK----- 56
Db 7198 WRVCLVKTGEEDHVLIVMHHIIYDGMNSVDILRGELQFYSAALRGQDPLHANPLIQY 7257
Qy 57 ----- 56
Db 7258 RDFAAWOREAKOVEEHORQLGYWSKQIVDSTPAELLTDLPRPSILSGRAGSVDDVTIEGSV 7317
Qy 57 ----- 61
Db 7318 YGALQSFCTRSTVTFWLLTVFRIAHFRLTAVDATICTPIANRRRPELETLVCCFVNT 7377
Qy 62 ----- 61
Db 7378 QCMRISIADDDNFEGLVRQVRNVATAAYANQDVPFERIVSALVPGSRNTRNPLVQLMFA 7437
Qy 62 ----- 61
Db 7438 VOSVEDYDQVRLEGLSVMPGEASTREDMEPHLVGDKLTGSLVYSSDLFEOGTTQNF 7497
Qy 62 ----- 61
Db 7498 VDIQFQCLRSVLDOPLTPISVLFPNSNAISNLESDDLEMTSDYPRORTVVDFREQAAI 7557
Qy 62 ----- 61
Db 7558 CPDSIAVKDSSQLTYAQLDEQSDRVAAMLHERHMPAESLVGLVSPRSCETIIAYFGINK 7617
Qy 62 ----- 61
Db 7618 ANLAYPLDVYAPDARLAAAILDVTVEGERLLLLGAGVPOPGIOI PRLSTAYIAEALSHATT 7677
Qy 62 ----- 65
Db 7678 VDVTSIQPSATSALAYVIFTSGTGKPKGVMIEHRGIVRLVDRDNTVNVPPSGSALPVSH 7737
Qy 66 ----- 65
Db 7738 FSNLAWDAATWEIYTAVLNGTGVVCI DRDTMLDIAALNSTFRKENVRAAFTPAFLKQCL 7797
Qy 66 --APNITANL----- 73
Db 7798 AETPELVANLEILHTAGDRLOPDGDANLAGKTAKGIFNVLGHTEHTAYSTFYPVVGEEFT 7857
Qy 74 ----- 73
Db 7858 VNGVPVGRGISHSHAYIIDRHOKLVPAGVMGELITGOGVARGYTDGALNKRFEVYIDIN 7917
Qy 74 ----- 73
Db 7918 GKSTWSYRTGDKARYRPRDGOLEFFGRMDQMKIRGVRIEPEGEVELTLDHKSVLAAATV 7977
Qy 74 ----- 73
Db 7978 VRRPENGPEMIAFITDAEDDVOTHKAIYKHLQGLPAYMIPSHLVILDOOMPVTNGKV 8037
Qy 74 ----- 73
Db 8038 DRKDALARAOQVQKRSSTAARVPRDEVEAVLCEEYSNLLEVEVGITDGFDLGCHSLA 8097
Qy 74 ----- 73
Db 8098 TKLAARLSROLNTRVSKVDQDQILADLADIIRGSHRHDPIPATPYTGPVEQSPAQR 8157
Qy 74 ----- 73
Db 8158 LWFLEQLNGASWYLPFAIRMRGPLOTKALAVNALVHRHEALRTTFEDHGVGVQVI 8217
Qy 74 ----- 73
Db 8218 QPKSSQDLRIIDLSDAVDDTAYLAALKREQTAFDLTSEPGHVRVSLRLGDDDDYILSIVM 8277
Qy 74 ----- 73
Db 8278 HHIIISDGTVDVLRLOELGOFYSAAIRGOBPLSOAKSLPIQYRDPFAVMQRQENQIKEAQAK 8337
Qy 74 ----- 73
Db 8338 LKYSQOLADSTPCEFLTDLPRLPSILSGEADAVPMVIDGTVVQLLTDFCRTHQVTSFVL 8397
Qy 74 ----- 73
Db 8398 LAAFRTAHVRLTGLTDATVGTPIANRRRPELEGLIGFFVNTQCMRAISETETFESLVQ 8457
Qy 74 ----- 73
Db 8458 VRLTTEAPANQDVPFEQIVSTLLPGSRDTSRNLPLVQVNFALQSOODLGRIGLEGMTDEA 8517
Qy 74 ----- 73
Db 8518 LETPLSTRDLLEVHLFOEBVGKLSGLLYSTDLFEVETIRGIVDFLETLRRGLEQPKORL 8577
Qy 74 ----- 73
Db 8578 NAMPITDGIKLRDOGLLTVAKPAYPRESSVIDLFROQVAAAPDAIAWDSSTLTLYADL 8637
```



```
QY 74 ----- 73
Db 8638 DQCSNLAHLWCORNPETLVAVPAPRSCLTIIVAFGLVKANLAYLPDVNAPARIEA 8697
QY 74 ----- 73
Db 8698 ILSAVPGHKLVLVOAHGPELGLTWADTELVOIDEALASSSGDHEQIHASGPTATSLAYV 8757
QY 74 ----- 73
Db 8758 MFTSGSTGPKGVMDHRSIIRLVKNSDVVATLPTVRMANVSNLAFDISVQEIYTALLN 8817
QY 74 ----- 73
Db 8818 GCTLVCLDYLTLDLSKILYNVFEAQNAAFTPVLLKQCLGNMPAIISRLSVLPNVGR 8877
QY 74 ----- 73
Db 8878 LDAHDAVASGLIODAVYNAYGPTENGQMSTMYKVDVNEPFGVPIGRSITNSGAYVMD 8937
QY 74 ----- 73
Db 8938 GNOOLVSPGVMEIVVTGDGLARGYTDSDALDEDRFVHVITIDGEENIKAYRTGDRVYRPK 8997
QY 74 ----- 73
Db 8998 DFEIEPGRMDQVKIRGRIEPAEVEHALLGHDLVHDAAVLVRKPAQOPEMIAFITSQ 9057
QY 74 ----- 73
Db 9058 EDETIEHESKQVCGMGEHFDVSRYADIKDLDTSTFGHDFLGWTSYDGVDPVNEKME 9117
QY 74 ----- 78
Db 9118 WLDETTASLLDNRPGGHILEIGAGTGMLSNLKGVDGLQYVGLDPAAPSAATFVNEAVKS 9177
QY 79 ----- 78
Db 9178 LPSLACKARVLVGTALDGLSKNEIQPELVVINSVAQYFTPTSEYLIKVKKAVVEVPSVK 9237
QY 79 ----- 78
Db 9238 RVFFGDIRSOALNRDPLAARAVRALGDNASKEQIREKIAELSEEEELLVDPAFVSURS 9297
QY 79 ----- 78
Db 9298 QLPNIKHVEVLPKLMKATNELSSYRYAAVLHISHNEEBQLLIQDIDPTAWVDFAATQKDS 9357
QY 79 -----SVC 81
Db 9358 OGLRNLQOGRDDVMIAVGNIPYSKTIIVERHIMNSLDQDHVNSLDGTSWISDARSAAAIC 9417
QY 82 ----- 81
Db 9418 TSFDAPALTQAKEGFRVELSWARQSONGALDAVFHRLATDANCERSVLVHFPTDQ 9477
QY 82 ----- 81
Db 9478 GRQLRTLNRPLQRAQSRIESQVFEALQTPALPAYMTPSRIIVLPQMPNANGKVDKQL 9537
QY 82 ----- 81
Db 9538 ARRAQVVAKRKAVSARVAPRNDTEIVLCEEYADILGTEVGTIDNFDMDGHSMLWATKLAA 9597
QY 82 ----- 81
Db 9598 RLSRRLDTRVTVEFDKPVKLADIAASIEQGSTPHLPFIASSVSYSGPVEQSYAQRLWFLD 9657
QY 82 ----- 81
Db 9658 QFNLNATWYHMSLAMRLGLPLNMDALDVALRALEQRHETLRTTFFBAQKDIGVOVVHEAGM 9717
QY 82 ----- 81

QY 9718 KRLKVLDSLDSKNEKHEMAVLENEQMRPFTLASPCWKGHRLARLGPTEYILLSVMHMFSD 9777
QY 82 GWS ----- 84
Db 9778 GNSVDILRQELGQFYSAALRGDRDPLSQVKPLPIQYRDFAAWQKEAAQVAEHEROLAYMEN 9837
QY 85 ----- 84
Db 9838 QLAOSTPCELLTDFRPQFLSGKAGVIPVTIEGVPVYEKLLKFSKERQVTLFSVLLTAFRA 9897
QY 85 ----- 84
Db 9898 THFRLTGAEDATICTPIANRMRPELEHIIIGFVNTOCMRLLLDTGSTFESLVQHVRSVAT 9957
QY 85 ----- 84
Db 9958 DAYSNQDIPFERIVSALLPGSRDASRPLIQLMFALHSQDPLNITLEGLEHERLPTSVA 10017
QY 85 -----QTIN----- 88
Db 10018 TRFDMEPHLFQEPNKLSGSILFADELFPQPETINSVVTVFOEILRRGLDQPOVSISTMPLT 10077
QY 89 ----- 88
Db 10078 DGLIDLEKLGLLBIESSNPPRDYSVVDFRQQVAANPNAPAVVDSETSMSYTSLDQKSEQ 10137
QY 89 -----PE----- 90
Db 10138 IAAWLHAQGLRPESLICVMAPRSFETIVSLFGILKAGAYVLPDVNSPAARTQPTLSEVE 10197
QY 91 ----- 90
Db 10198 GKRLVLLSGIDMPOSDRMDVETARIQDILTNTKVERSDPMGRPSATSLAYVIFTSGSTG 10257
QY 91 ----- 90
Db 10258 RPKGVMEHNRNLRVQKSNVTSQLPQDLRMAHINLAFDASIEWEIFTAILNGGALICID 10317
QY 91 ----- 90
Db 10318 YFTLLDSQALRTTFEKARVNATLFAPALLKECLNAPTFLFEDLVLYIGGDRLDATDAK 10377
QY 91 ----- 90
Db 10378 IQALVKGTVYNAYGPTENTVMSTIYRLTDGESYANGVPIGNAVSSSGAYIMDKORLVPP 10437
QY 91 ----- 90
Db 10438 GYMGELVVGDLARGYTNSTLNADR FVDIVINDQKARAYRTGDRTRVPRPKDGSIEPGR 10497
QY 91 ----- 90
Db 10498 MDQQVKIRGHRVEPAEVEQAMLGNAIHDAAVVVQAVDGOETEMIGFVMSADRSEEE 10557
QY 91 -----DDT----- 93
Db 10558 EITNQVQEWEDHFESTAVAGIEADQATLGRDFTSWYNGNLIIDKAEMBEWLDLDTMQS 10617
QY 94 ----- 93
Db 10618 LLDKEDARPCAEICTGTGMVLFNLPKNDGLESYVGIPEPSRSAALFVDKAAQDPGLQGT 10677
QY 94 ----- 93
Db 10678 QILVGTAEIDIKLVKDFHPDVVVVINSVAQYPPSRSYLVQIASELIHMTSVKTIFFGDMRSW 10737
QY 94 -----DP----- 95
Db 10738 ATNRDFLVSRALYTLGDKATKDQIROEVARLEENEDELLVDPAFTSLTSOWFGKVKHVE 10797
QY 96 ----- 95
```

```
Db 10798 ILPKRMTSNELSYRYAAVLHICRDBGRNRYGRVHVSVEENAWIDFASGMDRHALVQ 10857
Qy 96 ----- 95
Db 10858 MLDERRDAKTVAIGNIPHSNTINERHFTTSLDTEGEIAQDSLDGSAWQATKAAWARP 10917
Qy 96 ----- 95
Db 10918 CLSVTELVEIGQAAGFRVEWSAQRSQHGALDVVFHHEDDRVRGVLINFPPTDFERLPP 10977
Qy 96 ----- 95
Db 10978 STGLTSRPLQRIORRRFESQIREQLTLLPPYMPVSRIVLERMPLNANSKVORKELARK 11037
Qy 96 ----- 97
Db 11038 ARTLOTIKPSATRVAPRNDIEAVLUCDEFQAVLGVTGVMONFFELGCHSLMATKLAARLS 11097
Qy 98 ----- 101
Db 11098 RRLDTRVSKDIFNQPIQLDLADVVTQGSAPHEAIPSTPYSGPVEQSFSQGRWLWFLDOLN 11157
Qy 102 ----- 101
Db 11158 LNASWYHMLASRLRGLRIEALQSALATIEARHESLRTTFEBQDGPVQIVRAARNKQL 11217
Qy 102 ----- 101
Db 11218 RIIDVSGTEDAYLAALKQBQDAAPDLTAEPGRVALLRGLPDDHDHVLVSIVMHHIISDGWSV 11277
Qy 102 ----- 103
Db 11278 DILROELQLYSNASSQAPLPIQYRDFAIWQKDSQIAEHQKQLNWKVKQLVNSKPAEL 11337
Qy 104 ----- 103
Db 11338 LADTRPKALSGDADVIPIDDOVYQNLRSFCRRHVTSFVALLAAFRAAHYRLTGAED 11397
Qy 104 ----- 103
Db 11398 ATIGSPIANRRPELEGLICFVNTQCLRIPVKSEDTFTLVKQARETATEAODNODVPF 11457
Qy 104 ----- 111
Db 11458 ERIVSSMVASSRDTSRNPLVQVMFAVHSQHDGLNIRLEGVEGPKVMAASTRFAEMHLF 11517
Qy 112 ----- 111
Db 11518 EDQMLGNVVFVKDLFSESTIRSVAVFOETLRRGLANPHANLATPLTDGLPSLRSIC 11577
Qy 112 ----- 113
Db 11578 LQVNPQDYPDASVIDVFREQVASIPKSTAVIDAASSQLTYTELDERSSOLATWLRQVTV 11637
Qy 114 ----- 118
Db 11638 PEELVGLAPRSCETIIAFLGIKANLAYPLDVNAPAGRIETILSSLPGNRLILIGSDT 11697
Qy 119 ----- 119
Db 11698 QAVKLHANSVRFTRI3DALVESGSPTEELSTRPTAOSLAYVNFSTSGTGVPKGMVMEHR 11757
Qy 120 GVT----- 122
Db 11758 GITRLKVN5NNAKQAAAAIAHLSNIAFDASSWEIYAPLLNGGTVCIDYVTTIDIKAL 11817
Qy 123 ----- 122
Db 11818 EAVFKQHIRGAMLPPALLKQCLVSAPTMISSLEILFAAGDRLSSQDAILARRAVGSGVY 11877
Qy 123 ----- 122
Db 11878 NAYGPTENTVLSTIHNIGENAFSGVPIGNVNSGAFVMDQNOQLVSAGVIGELVVTG 11937
Qy 123 ----- 122
Db 11938 DGLARGYTSKLURVDRFIYITLDGNRVRAYRTGDRVRRHPKDGQIEFFGMRDQOIKIRGH 11997
Qy 123 ----- 122
Db 11998 RIEPAEVEQALARDPAISDSAVITQLTDEEPPELVAFPSLKGNAANGTNGVNGVSDQEKID 12057
Qy 123 ----- 122
Db 12058 GDEQHALLMENKIRHNLOALLPTYMIPSRIIHVDQLPVNANGKIDRNELAVRAQATPRTS 12117
Qy 123 ----- 122
Db 12118 SVSTVYAPRNDIETIICKEFADILSVRVGITDNFFDLGCHSLIATKLAARSLRDLTRVS 12177
Qy 123 ----- 122
Db 12178 VRDFTPTVVGOLAAASIOOGSTPHEAIPALSHSGPVQOSPAQGRLWFLDRFNLNAWYIM 12237
Qy 123 ----- 122
Db 12238 PFGVRLRGLRVDAQTALRALEERHELLRTTFEBQDVGVMQIVHSPRMDICVVVDISGA 12297
Qy 123 ----- 122
Db 12298 NEDLAKLKEEQOAPPNLSTEVAMRVALPKAGENHHILSIVMHHIISDGWSVDIIFOEBLQ 12357
Qy 123 ----- 122
Db 12358 FYSVAVRGHDPLSQVKPLPIHYRDFAVWQRODKQAVHVESQLOQWIEQLADSTPAEILSD 12417
Qy 123 ----- 122
Db 12418 FNRPEVLSGEAGTVPIVIEDEVYKLSLFCRNHQVTSFVLLAARVAHYRLTGAEDATI 12477
Qy 123 ----- 122
Db 12478 GTPIANRNRPELEDLIGFFVNTQCMRIABEHDNPLSVVRRVRSTAASAFENQDVPFERL 12537
Qy 123 ----- 122
Db 12538 VSALLPGSRDASRNPLVOLMFVHVSQRNLKQLEGEGETPYTATTTRPDVEFHLPQD 12597
Qy 123 ----- 122
Db 12598 KGLAGNVVFAADLFEAATIRSVVEVFHEILRRGLDQPDIAISTMPLVDGLAALNSRNLP 12657
Qy 123 ----- 122
Db 12658 VEDIEPDPAEASVVDFQTVVQVNPALAVTDTSTKLYAELQOQSDHVAAMLSKQKLP 12717
Qy 123 ----- 122
Db 12718 AESIVVVLAPRSSETIVACIGILKANLAYLPMDSNVPEARQAILSEIPGEKVVLLGAGV 12777
Qy 123 ----- 122
Db 12778 PIPDNKTADVMMVFISDIVASKTDKSYSGTPRPSASSLAYVIFTSGSTGRPKGMVMEHRG 12837
Qy 123 ----- 122
Db 12838 VISLVKQNSRIPQSLRMAHVSNLAPDASVWEIFTTLLNGGTLCISYFTVLDKALSAA 12897
Qy 123 ----- 122
Db 12898 FSDHRINITLLPALLKQCLADAPSVLSLESLYIGGDRLDGADATKVKDLVKGKAYNAY 12957
Qy 123 ----- 122
Db 12958 GPTENSVMSTIYTIETHEHETFANGVPIGTSLGPKSKAYIMDQDQQLVPAGVMGELVYAGDGL 13017
```

QY 123 ----- 122
Db 13018 ARGYDPSLNTGRFNIHITIDCKQVAYRTGDRVRYRPRDYQIEFFGRLDQOIIRGHRIE 13077
QY 123 ----- 122
Db 13078 PAVEQALLSDSSINDAUVVSAQNKEGLEMGVYITTTAAQSDVKKEASKNVQWEAHFDS 13137
QY 123 ----- 122
Db 13138 TAVANIGGIDRDALGODFLSWTSMYDGLIPREMOEWLNDTWRSLLDNQPPGKVLEIGT 13197
QY 123 ----- 122
Db 13198 GTGMVFLNCKVEGLQSYAGLEPSRVTAWNKAIEFTFSLAGSARVHVGTAEIDISSIDG 13257
QY 123 ----- 122
Db 13258 LRSDLVINSVAQYFPSREYLAELTANLIRLPGVKRIFFGDMRTYATNKKDLVARAVHTL 13317
QY 123 ----- 122
Db 13318 GSNASKAMVROQVAKLEDDEEELLVDPAFFTSLSDOFPDEIKHVEILPKRMAATNELSSY 13377
QY 123 ----- 122
Db 13378 RYAAVIHVGHQMPNGEDEDKQAVKIDINPKAWVDFAGTRMDRQALLQLLQDRQRDDVV 13437
QY 123 ----- 122
Db 13438 AVSNIPYSKTIMERHLSQSLDDEGTSVADGTAMISRTQSAKECPALSVADLIEIGK 13497
QY 123 ----- 122
Db 13498 IGFEVEASWARQSORGLDAVFRFEPHSHGVNFRPTEHKGRSSSLTNRPLHLLO 13557
QY 123 ----- 122
Db 13558 SRRLKVRERLOSLPPYMIPSRITLLDQMLPTNGVKVDRKKLARQARVIPRSAASTLD 13617
QY 123 ----- 126
Db 13618 FVAPRTEIENVLCEBFTDLLGVKVGITDNFFELGGHSLATKLSARLSRRLDAGITVKQV 13677
QY 127 ----- 126
Db 13678 FDOPLADLAASILQSSSRHSRPSLPYEGPVEQSAQRLWFLDQFNIDALWVLPAL 13737
QY 127 ----- 126
Db 13738 RMRGPLQVDALAAALVALERHESLRTTTEERDGVGIQVQPLRTTKDIRIIDVSGMRDD 13797
QY 127 ----- 132
Db 13798 DAYLEPLOKEQTPFDLASEFCWRVALLKLGKDDHILSIVMHHIISDQWTEVLORELQ 13857
QY 133 ----- 132
Db 13858 FYLAASKGAPLSQVAPLPIQYRDFAVWQOEQVAESORQDYWKQLADSPAEILLAD 13917
QY 133 ----- 132
Db 13918 YTRPNLSGEAGSVSFVINDSVYKLSVFCRSQVTTFTTLAAPRAAHYMTGSDDATI 13977
QY 133 ----- 136
Db 13978 GTPIANRNPENLIGCFVNTQCMRITIGDETFSLVQVRSRTTATAPENQVPPERI 14037
QY 137 ----- 144
Db 14038 VSTLSAGSRDTSRNPVLQVLLFAVHSQQGLGRIGLDGVVDEPVLSTVSTRFDLPHAFQEA 14097
QY 145 ----- 144

Db 14098 DRLNGSMVFATDLPQETIQGFVAVVEVLORGLEQPOSPIATPLAEGIAQLRDAGALO 14157
QY 145 ----- 144
Db 14158 MPKSDYPRNASLVDVFQOQAMASPTVAVTDSKLTYAELDRSLDQAASYLRROQLPAE 14217
QY 145 ----- 144
Db 14218 TWAVLAPRSCETIIAFLAILKANLAYMPLDVNTPSARMEAIISVPGRRLLILVSGVRH 14277
QY 145 ----- 144
Db 14278 ADINPVNAKTMILSDTVTGTDAIGTPPEPLVVRPSATSLAYVIFTSGTGKPKGMVVEHRA 14337
QY 145 ----- 144
Db 14338 IMRLVKDSNVVTHMPPTARMHVNTIAFDVSLFEMCATLLNGGTLVCIDYLLDSTMLR 14397
QY 145 ----- 144
Db 14398 ETFEREQVRAAIFPPALLRQCLVNMPPDAIGMLEAVYVAGDRFHSRDARATQALAGPRVYN 14457
QY 145 ----- 144
Db 14458 AYGPTENAILSTIYNIDKHDPYNGVPIGSAVNSGAYVMDRNOQLLPPGMVGLVVTGE 14517
QY 145 ----- 144
Db 14518 GVARGYTDASLDTRFVTVTIDGORQRAYRTGDRVRYRPGFQIEFFRLDQOAKIRGHR 14577
QY 145 ----- 144
Db 14578 VELGEVEHALLSENSVTDAAVVLRTMEEEDPOLVAFVTTDHEYRSGSSNEEDPYATQAA 14637
QY 145 ----- 144
Db 14638 GDMRKRLRSLPYVMVPSRVITLROMPLNANGKVDKDLARRAQMTPTASSSGPVHVAPR 14697
QY 145 -----GVT 147
Db 14698 NETEAAICDEFETILGVKVGIT 14719

RESULT 2

US-09-369-364A-7
; Sequence 7, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
US-09-369-364A-7

Query Match 30.7%; Score 237; DB 4; Length 997;
Best Local Similarity 18.2%; Pred. No. 0.00038;
Matches 61; Conservative 24; Mismatches 62; Indels 188; Gaps 6;

QY 1 AAGGILHLEL-----LVAVG----- 15
Db 53 AGGSFLSYELWPRALKRVDVRRDAPAFYLOYRGRFLRFLNTANQHLLAPGCVSETRR 112
QY 16 -----PD-----VFQAHQED----- 25

Db 113 RGLGRAHRAHTPACHLLGEVDPELEGLAAISACDGLKGVFQLSNEDYFIEPLDSAP 172
Qy 26 -----
Db 173 ARPGAHPVVKYKQAPERLAQRGDSSAPSTCGVQVYPELESRRERWEQRCQWRRLRR 232
Qy 26 -----
Db 233 LHORSVSEKWCETLVADAKMVEYHGOQVESYVLTIMNVAGLFHDPISGNPIHITIV 292
Qy 56 KXVILTEPEGAPNITANITSSLLSYCGMSQINPEDDTPDGHADLVLYITRFDLELPGN 115
Db 293 RLVLLEDEEDLKITHADNTLKSFCWKQKSINKMGDAHPLHHTAILLTRKDL-CAAMN 351
Qy 116 R-QVRGVTOLGGACSPWSCLI7EDTGFDLGVTI 148
Db 352 RPECETLGLSHVAGMCQPHRSCSINEDTGLPLAFTV 386

RESULT 3

US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin.
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-6

Query Match 30.4%; Score 235; DB 4; Length 11877;
Best Local Similarity 1.1%; Pred. No. 0.33;
Matches 91; Conservative 21; Mismatches 36; Indels 8329; Gaps 25;

Qy 1 AAGG-----
Db 450 AAGQAVTDVPADRNAGDFYDPRSAPGRNSRWGFIEDVDRFDAAPFGISPREAEM 509
Qy 5 -----
Db 510 DPOQLALELGEALERAGIDPSSLTGTTRGVFAGAIMDDYATLKHROGGAALTHTVTG 569
Qy 5 -----
Db 570 LHRGIIANRLSYTLGLRGPMVVDGQSSSLVAHVILACESLRGESELALAGVSLNLP 629
Qy 5 -----
Db 630 DSIIGASKFGLSPDGRAYTFDARANGYVRGEGGFVWLKRLSRAVADGDPVLAVIRGSA 689
Qy 5 -----
Db 690 VNNGGAAQGMTPDQAQEAVALREAHERAGTAPADRVYVELHGTGTGTPGDPIEAAALGAA 749
Qy 5 -----
Db 750 LGTRPAGQPLLVGSKVTNIGHLEGAAGIAGLIKAVLAVRGALPASLNYETPNPAIPFE 809
Qy 10 -----
Db 810 EMLNRVNTPELPEPHDQRMVGVSSFGMGGTNAHVLEAPGCGRCASIVESTVGS 869

Qy 10 -----
Db 870 AVGGVVPVWVSAKSAALDAQIERLAAPASRDRTDGVDAVDAVDAVAVVLGG 929
Qy 10 -----
Db 930 RAQFEHRAVVVGSGPDDLAALAAPEGLVRGVASGVRVAFVFPGGGTOMAGKAEELDS 989
Qy 18 --VFQA-----
Db 990 SAVFAAAMAECEAALSPPYVDSLEAVVRQAPGAPTLERVDVVPVTFFAVMVSLARVWQH 1049
Qy 22 -----
Db 1050 GVTPOAVVGHSGOIEAAAYVAGALSDDAARVVTLRSKSIAAHLAKGKGMLSLASEDAV 1109
Qy 22 -----
Db 1110 LERLAGFDGLSVAAVNGPTATVVGDPVQIEELARACEADGVRARVIVPDYASHSROVEI 1169
Qy 22 -----
Db 1170 IESELAEVLAGLSPOAPRVPPFSTLEGAWITPVLDDGGVYRNLRHVRGFAPAVETLATD 1229
Qy 22 -----
Db 1230 EGFTHFEVSAHPVLTMALPGTVTGLATLRDNGQDRLVASAEAWANGLAVDWSPLLP 1289
Qy 22 ----HQED-----TER-----
Db 1290 SATGHHSDLPTYAFOTERHMLGEIBALAPAGEPAVQPAVLTEAAEPAELORDEOLRVIL 1349
Qy 29 -----
Db 1350 DKVRAQTAQVLGYATGGQIEVDRTFREAGCTSLTGVDLRRNRINAFAFVGRMAPSMIFDPPT 1409
Qy 29 -----
Db 1410 PEALAEQLLLVVHGEAAANPAGAPAPVAAAGAVDEPVAIVGMACRLPGGVASPEDLWRL 1469
Qy 29 -----
Db 1470 VAGGDAISEFPQDRGMDVEGLYHPDPEHPCTSYVROGGFTENVAGFADAAFFGISPREAL 1529
Qy 29 -----
Db 1530 AMDPOORLLLETSSWEAVEDAGIDPTSLRGROGVFTGAMTHEYGPSLRDGGEGLDGYLLT 1589
Qy 33 N-----LNIGAEALLR-----
Db 1590 GNTASVMSGRVSYTLGLEGPALTYDTACSSSLVALHVAQALRKGEVDMALAGGVAVMPT 1649
Qy 43 -----
Db 1650 PGMEFEFSRORLAGDGRSKAFASADGTSWSEGVVLLVERLSDARRNGHOVLAVVRGS 1709
Qy 43 -----
Db 1710 ALNODGASNLTPNGPSQORVIRRALADARLTSTDVDVVEAHGTGTRLGDPIDEAQALIA 1769
Qy 43 -----
Db 1770 TYGGRDDEQPLRLGSLKSNIGHTQAAAAGVSGVIKVMQMRHGLPKTLHVDEPSDOIDW 1829
Qy 43 -----
Db 1830 SAGAVELLTEAVDWPEKQDGLRRAAVSFGISGTNAHVLEAPVVEGASVVEPSVGS 1889
Qy 48 -----
Db 1890 SAVGGVTPVWVSAKSAALDAQIERLAAPASRDRTDDADAGAVDAVAVHLADGQAQF 1949
Qy 51 -----

```
Db 1950 EHRAVALGAGADLVQALADPDGLIRGTASGVGRVAFPPGGGTQWAGMGAEILLDSSAVF 2009
Qy 51 -----RVHLV----- 55
Db 2010 AAMAECEAALSPYVDWLEAVVROAPGAPTLEKVDVQPVTFVWVSLARVWQHGVTP 2069
Qy 56 ----- 55
Db 2070 QAVVGHSGEIAAAVAGALPLDDAARVVTLRSKSIAAHLACKGGMGLSALNEDAVLERL 2129
Qy 56 ----- 55
Db 2130 SDFDGLSVAAVNGPTATVSGDPVQIEELAQACKAGFRARIIPVDYASHSRQVEIIESE 2189
Qy 56 ----- 55
Db 2190 LAQVLAGLSQAPRVPFFSTLEGTWITEPVLDTGYWYRNLHRVGFAPAIETLAVDEGFT 2249
Qy 56 ----- 55
Db 2250 HFVEVSAHPVLTMTLPETVGLTLLRREGGQGERLVTSIAEAWVNGLPVAMTSLLPATAS 2309
Qy 56 ----- 55
Db 2310 RPGLTYAFOAERYWLENTPAALATGDDWRYRIDWKRLPAEAGSERTGLSGRWLAVTPED 2369
Qy 56 ----- 55
Db 2370 HSAQAAVLTALVDAGAKVEVLTAGDDDDREALAARLTALTGTGDTGVVSLDGLVPQV 2429
Qy 56 -----KMVILTEPE----- 64
Db 2430 AWOALGDAGIKAPLWSVTQGAHSVGRLDTPADPDRAMLWGLGRVVALEHPERWAGLVDL 2489
Qy 65 ----- 64
Db 2490 PAQPDAAALAHVLTALSGATGEDQIAIRTTGLHARRLARAPLHGRRPRTRDQPHGTVLIT 2549
Qy 65 -----GAPNITANLTSS----- 76
Db 2550 CGTGALGSHAARWMAHGAEHLLVSRSGEQAPGATQTAEILTASGARVTTAACOVADPH 2609
Qy 77 ----- 76
Db 2610 AMRTLDDAI PAETPLTA VVHTAGALDDGIVDTLTAEQVRAHRAKAVCASVLDLDRDLD 2669
Qy 77 ----- 76
Db 2670 LDAFVLFSSVSTLIGIPQGNYPHNAVYLDALAAARRRATGRSAVSAMGPMWGGMAAGD 2729
Qy 77 -----LLSVCGWSQTI----- 87
Db 2730 GVAERLRNHGVPQMDPELAALESALGRDETATVADIDWDRFYLAYSSGRPQPLVEEL 2789
Qy 88 ----- 87
Db 2790 PEVRRIIDARDSATSGGSSAOGANPLAERLAAAPGERTTEILLGLVRAQAAVLRMRS 2849
Qy 88 ----- 87
Db 2850 PEDVAADRAFKDIGDSLAVELNRNLTATGLQLPATLPDHPDTPALVSLRSEFLGD 2909
Qy 88 -----NPED----- 91
Db 2910 EETADARRSAALPATVAGAGAGAGACTDADDPIAVMSCRYPGDIRSPEDLWRMLSEGG 2969
Qy 92 -----DTDP----- 95
Db 2970 EGITFPFTRCWDLDGLYDADPDALGRAYVREGGLHDAAEFDAEFFGVSPREALANDPQ 3029
Qy 96 ----- 95
```

```
Db 3030 QRMILLTSWEAFERAGIEPASLRSGSTCGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPS 3089
Qy 96 ----- 95
Db 3090 VASGRIAYTFGLEGPATTVDTCSSSLTALHLAVRALRSGBCTMALAGVAMMATPHMPV 3149
Qy 96 ----- 95
Db 3150 EFSQRALAPDGRSKAFSADADGFGAAEGVGLLLVERLSDARRNGHPVLAUVVRCGTAVNQD 3209
Qy 96 ----- 95
Db 3210 GASNGLTAPNGFSQORVIRQALADARLAPGDDIDAVETHGTGTSIGDPIEAOQLOATYQKE 3269
Qy 96 ----- 95
Db 3270 RPAERPLAIGSVKSNIGHTQAAAGAAGIIMVLAHRHGTLPKTLHADEPSPHVDWANSGL 3329
Qy 96 ----- 95
Db 3330 ALVTEPIDWPAGTGPRAAVSSFGISGTNAHVVLEQAAPDAAGEVLGADEVPEVSETVAMA 3389
Qy 96 ----- 95
Db 3390 GTAGTSEVAEGSEASEAPAAPGSREASLPGLHPMWLSAKDRQSLRGQAAALHAWLSEPA 3449
Qy 96 ----- 95
Db 3450 DLSADGPARLURDVGVTLTATSRTPAFHRAAAVTAADRDGFLDGLATLAOGGTSAHVHLDTA 3509
Qy 96 -----GHADL----- 100
Db 3510 RDGTTAFLTGTGQSQRPGAGRELYDRHPVFAHALDEICAHLDGHLLEPLLDVMPAAEGSA 3569
Qy 101 ----- 100
Db 3570 EAALLDETRYQCALFALEVALFRLVBSWGMRPAAALGHVSGETAAAAHVAGVFSLADAAR 3629
Qy 101 ----- 100
Db 3630 LVAARGRLMOELPAGGAMLAVOAAEDEIRVWLETEERYAGRLOVAAVNGPEAAVLSGDAD 3689
Qy 101 ----- 100
Db 3690 AAREAEAYWGLGRRTRALRVSHAFHSAHMDGMLDGFRVAVLETFEPRRPSLTVVSNVTGL 3749
Qy 101 ----- 100
Db 3750 AAGPDDLCDPEYVWRHVHRTVRFLDGVRVLRDLGVRTCLELGPDPGLVTAMAADGLADTPA 3809
Qy 101 ----- 100
Db 3810 DSAAGSPVGPAGSPADSAAGALRPRPLLVALLRKRKSETETVADALGRAHAHGTGPDWH 3869
Qy 101 ----- 100
Db 3870 AWFAGSAHRVDLPTYSFRDRYWLDAADTAVDTAGLGTADHPLLGAVVSLPDRDG 3929
Qy 101 ----- 100
Db 3930 LLITGRLSLRTHPWLADHAVLGSVLLPGAMVELAAHAAESAGLRDVRLETLLEPLVLPE 3989
Qy 101 ----- 100
Db 3990 HGOVELRVTVGAPAGEPGGESAGDARPVSLHSRLADAPAGTAWSCHATGLLATDRPELP 4049
Qy 101 ----- 100
Db 4050 VAPDRAAMWPPQGAEEVPLDGLYERLDGNGCLAFGLPFOGLNAVRYEVEGFADIALPATT 4109
Qy 101 ----- 100
Db 4110 NATAPATANGGSGAAAAAPYGIHPALLDASLIHAIYAVGLVDEPELVRVPFHWSGVTVHAAG 4169
```

Qy	101	-----	100	Qy	121	-----	120
Db	4170	AAAARVRLASAGTDAVSLTDEGRPLVSVRLTLRPVTADQAAASRVGGLMHRVAWRP	4229	Db	5250	AVNQDASSGLTAPNGPSQORVIRRALADARLTTSDDVVVEAHGTGTRLGDPLEAQAVTA	5309
Qy	101	-----	100	Qy	121	-----	120
Db	4230	YALASSGEODPHATSYGPTAVLGKDELKVAALLESAGVEVGLYPDLLAALSQDVAAAGAPAP	4289	Db	5310	TYGQRDGEQPIRLGSLKSNIGHTOAAAAGVSGVIKVMQMRHGVLPKTLHVEKPTDQVDM	5369
Qy	101	-----	106	Qy	121	-----	120
Db	4290	RTVLAPLPAGPDGGAEGVRGTAVRTLELLQWLADENHLAGTRLLLVTRGAVRDPESGA	4349	Db	5370	SAGAVELLTEAMDPDKDGGGLRRAAVSFGVSGTNAHVULEEAPAAEBETPASEATPAVE	5429
Qy	107	-----	106	Qy	121	-----	120
Db	4350	DGGEDLSHAAANGLVRTAOTENPGRGLLDLADASSYRTLPSVLSDAGLRDEPOLALH	4409	Db	5430	PSVGAGLVFWLVSAKTPAALDAQIGRLAASFASQGRDAADPGAVARVLAGGRAEPEHRAV	5489
Qy	107	-----	106	Qy	121	-----	120
Db	4410	DGTIRLARLASVRPETCTAAPALAPECTVLLTGGTGGGLGLVARHVVGWVRRLLVSR	4469	Db	5490	VLGTQDDPFAQALTAPEGLIRGTPSDVGRVAFVFPQOGTQWAGMGAELLDSVKEFAAMA	5549
Qy	107	-----	106	Qy	121	-----	120
Db	4470	RGTAPGABELVHEALGADVSVAACDVADREALTAVLDAIPAHEPLTAVVHTAGVLS	4529	Db	5550	ECESALSRVDSLEAVVRQAPGAPTLERVDVQVPTFAVMVSLAKVMQHHGVTPOAVVG	5609
Qy	107	-----	106	Qy	121	-----	120
Db	4530	GTLPSTMTDEHVLRPKVDAAFLDLBELTSTPAYDLAAFWFSSAAAVFGAGQAYAAA	4589	Db	5610	HSQGEIAAAVAGALTLDAAARVVTLRSKSIAAHLGKGMISLALSEATRQRIENLHG	5669
Qy	107	-----	106	Qy	121	-----	120
Db	4590	NATLDALWRRRAAGLPALSLGHLGWLWETSMTGELGQADLRKMSRAGIGGISDAEGIAL	4649	Db	5670	LSIAAVNGPTATVVGSDPTQIQLAQACEADGVRRARIIPVDYASHAHVETIESLAEVL	5729
Qy	107	-----	106	Qy	121	-----	120
Db	4650	LDAALRRDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRRPSSAASATTA	4709	Db	5730	AGLSPTPEVPFFSTLEGAMITEPVLDTGYWYRNLHRVGFAPAVETLATOGETHFTEV	5789
Qy	107	-----	106	Qy	121	-----	125
Db	4710	GTAAGTGTADGAAETAATVLAADRAATVDGPORORLLELFFVVGVEAEVLGHARGHRTDAER	4769	Db	5790	SAHPVLTMTLPETVTGLTGLRREGQGOBERLVTSLAEAWNTGLTIWAPVLTATGHHPEL	5849
Qy	107	-----	114	Qy	126	-----	125
Db	4770	GFLLDGFSLTAVELNRNLSAGGLALPATLVFDHPSPAALASHLDAELPRGASDQDGAG	4829	Db	5850	PTYAFORRHVHDSPAVQSVQDSWRYRIDWKRLAVADASERAGLSGRWLVVVPEDRSA	5909
Qy	115	-----	114	Qy	126	-----	125
Db	4830	NRNGNNGTTASRSTAETDALLAQLTRLEGALVLTGLSDAPGSEEVLEHLRLSRSNVTGE	4889	Db	5910	EAPVLAALSGAGADPVQDVSPLGDRQRLAATLGEALAAAGGAVDGVLSLLAWDESAHP	5969
Qy	115	-----	114	Qy	126	-----	125
Db	4890	TGTGTASGAPDGAGCAEDRPHWAGDAGGSGEDGACVPDFMNASAEELFGLLDQDPSTD	4949	Db	5970	GHPAPFTRGTGATLTLVOALEDAGVAAPLMCVTHGAVSVGRADHVTSPAQAMVMGMRVA	6029
Qy	115	-----	114	Qy	126	-----	126
Db	4950	MSTVNEEKYLDYLRATADLHEARGRLRELEAKAGPVAIVGMACRLPGGVASPEDLWRL	5009	Db	6030	ALERPERWGGILDLPSDADRAALDRMTTVLAGGTGEDQVAVRASGLLARLVRASLPAGH	6089
Qy	115	-----	114	Qy	127	ACSPTW	132
Db	5010	VAGGEDAISEFPQDRGWDVEGLYDPNPEATGKSYAREAGFLYEAGEFDADFFGISPREAL	5069	Db	6090	TASPMWQADGTVLVTGAEPAAEAARRLRDAGHLLHLLTTPSGSEGAECTSGAEDSG	6149
Qy	115	-----	114	Qy	133	-----	132
Db	5070	AMDPOORLLLESWEAFEHAGIPAAATAGTSGVFTGVVMYHDYATRLTDVPEGIGYLG	5129	Db	6150	LAGLVASLADLGATATVVTCDLTDAEAAARLLAGVSDAHPLSAVLHLPPTVVDSEPLAATD	6209
Qy	115	-----	114	Qy	133	-----	132
Db	5130	GNSGSVASGRVAYTLGLEGPVAVTDTACSSSLVALHLVAQALRKGEVDMALAGGVTVMST	5189	Db	6210	ADALARVVTAKATAALHLDRLLREAAAAAGRPVVLVLFSSVAATWGGAGGAGYAAAGTAF	6269
Qy	115	-----	120	Qy	133	-----	132
Db	5190	PSTFVEFSRQRLAPDGRSKSPSSSTADGTSWSEGVGVLLVERLSDARRKGRHILAVVRGT	5249	Db	6270	DALAGQHRADGPTVTSVAVSPWEGSRVTEGATGERLRLGLRPLAPATALTALDTALGHG	6329
				Qy	133	-----	132

```
Db 6330 DTAVTIADVMSFAPGFTTARPGTLLADLPEARRALDEQOSTTAADTVLSRELGALTG 6389
Qy 133 ----- 132
Db 6390 AEOQRKQELVREHLAVLNHPSPBAVDTCRAPRDLGFDLSLTAVELNRNLKNATGUALPA 6449
Qy 133 ----- 132
Db 6450 TLVFDYPTPTLAEFLLEILGEQAGAGEQLPVDGGVDDEPVAIVGMACRLPGGVASPED 6509
Qy 133 ----- 132
Db 6510 LMRLVAGGEDAISGFPQDRGWDVEGLYDPDPDASGRTYCRAGGFLDEAGEFDADFISG 6569
Qy 133 ----- 132
Db 6570 REALANDPOORLLETSSWEAVEDAGIDPTSLGOQOVGVFAGTNGPHYEPDLLRNTAEDLEG 6629
Qy 133 ----- 132
Db 6630 YVGTGNAASIMSGRVSYTLGLEPAVTVDTACSSSLVALHLAVQALRKGECCGLALAGGVT 6689
Qy 133 ----- 132
Db 6690 VMSTPTTFVFSRQRGLAEDGRSKAPASADGFGPAEYGVMLLVERLSDARRNGHRVLAV 6749
Qy 133 ----- 132
Db 6750 VRGSVNDQASNGLTAPNGPSQORVIRRALADARLTADVDVVEAHGTGTRLGDPIDIAQ 6809
Qy 133 ----- 132
Db 6810 ALIATYGGORDTEQPLRLGSLKSNIGHTQAAAGVGIIKWQAMRHGVLPKTLHVDRPDS 6869
Qy 133 ----- 132
Db 6870 QIDWSAGTVELLTEAMDMWRKOEGLRRAVSSFGISGTNAHIVLEEAAPVDEPADPEPS 6929
Qy 133 ----- 132
Db 6930 VGVVYVPLVSAKTPAALDAQIGRLAAAFASQGRDADDPGAVARVLAGRAQFEHRAVALG 6989
Qy 133 ----- 132
Db 6990 TQDDLAALAAPEGLVRGVASGVRAVFVPGOGTQWAGMAELLDVSKFAAAMAECE 7049
Qy 133 ----- 132
Db 7050 AALAPYVDSLEAVVRQAPGAPTLERVDVQVPTFAVMVSLAKVMQHGVTPQAVVGHQS 7109
Qy 133 ----- 132
Db 7110 GEIRAAAYAGALSDDAARVVTLSKSIKAHLAGOGGMLSLSEAAVVERLAGFDGLSV 7169
Qy 133 ----- 132
Db 7170 AAVNGPTATVVSGDPTQIQELAQACEADGVRAIIPVDYASHSAHVETIESELADVLAGL 7229
Qy 133 ----- 132
Db 7230 SPOTPOVPFFSTLEGAWITEPALDGGYWYRNLHRVGPAPAVETLATDEGTFHEVSAH 7289
Qy 133 ----- 132
Db 7290 PVLTMALPETVTGLTLRRDNGQHRLTTSIAEAWANGLTVDMWASLPTTTTHPDLPYIA 7349
Qy 133 ----- 137
Db 7350 FOTERYWPQDLSAAGDITSAGLGAHEPLLCAAVALADSDGCLLTGSLRTHPWLADH 7409
Qy 138 ----- 137
```

```
Db 7410 AVAGTVLLPGTAPVELAFRAGDOVGCOLVBEELTLDAPLVLPRRGAVRVOLSVGASDESGR 7469
Qy 138 ----- 137
Db 7470 RTFGLYAHPEADPGEAIEWTRHATGVLAARDRTAPVADPEAMPPEGASPEVDVGLYERFA 7529
Qy 138 ----- 137
Db 7530 ANGYGYGFLFGVRGVWRRGDEVFADVALPAEVAGAGARFGLHPALLDAVQAAGACRG 7589
Qy 138 ----- 137
Db 7590 VRRGHAAVRLRLDLLYAVGATALRVLPAGPDTVSYSAADSSGQPVFAADSLTVLPVD 7649
Qy 138 ----- 137
Db 7650 PAQLAAYSDDPTLDALHLLLEWTANDGAAQALPGAVVLCGGDADGLAAALRAGGTEVLSPDDL 7709
Qy 138 ----- 137
Db 7710 TDLVEAVDRGETPAPATVTVACPAAGPDGPEHVREALHGLSLALMQAWLADERFTDGRLLV 7769
Qy 138 ----- 137
Db 7770 VTRDAVAARSGDGLURSTGQAAVMGLGRSAQTESPGRFVLLDLAGEARTAGDATAGDLTT 7829
Qy 138 ----- 137
Db 7830 GDATVGGTSGDAALGSALATALGSGEPQALRQCALLVPRLARAAAPAAADGLAANDGLA 7889
Qy 138 ----- 137
Db 7890 ALPLPAALWRLPEPGTSGLESILTAAPGDAETLAPEPLPGQGVRIAIRATGLNFRDVL 7949
Qy 138 ----- 137
Db 7950 ALGMYDPALMGTEGAGVVTATGPGVTHLAPGRVDMGLLSGAYAPVVVADARTVARMEG 8009
Qy 138 ----- 137
Db 8010 WTPAOGASVPVVFILTAVYALRDLADVKPGERLLVHSAAGVGMAAVALARHWGVVHGTA 8069
Qy 138 ----- 137
Db 8070 SHGKWDALRALGLDDAHIASSRTLDPEAPRAASGGAGMDVWENSLAREPVDASRLG 8129
Qy 138 ----- 137
Db 8130 GGRFVEMGKTDVRDAERVAADHPGVGYRAPFDLGEAGPERIGEMLAEVIALFEDGVLRLP 8189
Qy 138 ----- 137
Db 8190 VTTWVRRARDAPRHVSQARHTCKVVTMPSGLDPEGTLLTGGTGALGGIVARHVVGW 8249
Qy 138 ----- 137
Db 8250 GVRRLLVSRGRTDAPGAGELVHELEALGADVSAACDVADREALTAVLDSIPAHEPLTA 8309
Qy 138 ----- 137
Db 8310 VHVTAGVLSGTLPSMTAEDVEHVLRPKVDAAPFLLEDLTSTPGYDLAAVFMFSSAAAVFG 8369
Qy 138 ----- 137
Db 8370 GAGQYAAANATLDALAWRRRTAGLPALSLGHLWAETSGMTGGGLSDTDRSLARSGAT 8429
Qy 138 ----- 137
Db 8430 PMDSELTLSLLDAAMRRDPALVPIALDVAALRAQORDGMLAPLLSGLTRSGRVGGAPVN 8489
Qy 138 ----- 139
Db 8490 QRRAAAGGAGEADTDLGGRLAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRD 8549
```

QY 140 TGFDL-----144
Db 8550 TGFDSLTVNELNRNLNAATGLRLPATLVDPHPTPGELAGHLDELATAAGGWAEGTSG 8609
QY 145 -----144
Db 8610 DTASATDRQTAAELDRLEGVLASLAPAAAGRPPELAARLRALAAALGDDGDDATDLDE 8669
QY 145 -----144
Db 8670 ASDDLFSFDKELGSDFMANNEDKRLDYLRVTAELQONTLRREIEGRTHEPVAIVG 8729
QY 145 -----144
Db 8730 MACRLPGGVASPEDLMQLVAGDGAISEPPQDRGNDVEGLYDPPDASGRTYCRSGFLH 8789
QY 145 -----144
Db 8790 DAGEFDADFGISPREALAMDPOORLSLTTTAWEAIESAGIDPTALKGSLGVFVGWHTG 8849
QY 145 -----144
Db 8850 YTSQTTAVQSPEGLHVSAAALGFLSGRIAYVLGTGDPALTVDTACSSSLVALHLAVQ 8909
QY 145 -----GVTI 148
Db 8910 ALRKGCDMALAGGVTV 8926

RESULT 4
US-09-036-987A-6
; Sequence 6, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow AgroSciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5588 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-036-987A-6
Query Match 27.9%; Score 216; DB 4; Length 5588;
Best Local Similarity 1.7%; Pred. NO. 0.6;
Matches 92; Conservative 25; Mismatches 29; Indels 5123; Gaps 30;
QY 2 AGGI-----5
Db 230 AGGISMATPGAFVFSRQALASDGRCKPPADAADGTGCGAGMLLLERLSDDARRNH 289
QY 6 -----5
Db 290 PVLAAVGSAINQDGTSGNLTAPSGAQORVIRQALANAGLSPAEVDVVVEAHGCTGALGD 349
QY 6 -----LHL 8
Db 350 PIEAQUALIATYGANRSADHPLLLSLKSNIGHTQAAAGVAGVIKSLAIRHREMPRLHI 409
QY 9 -----8
Db 410 DOPSHVMSAGAVRLTDSVDMPLGRPRRAGVSFGMSGTNAHLIIVEVDEPVGST 469
QY 9 -----8
Db 470 EPTGAPWPLSGKTETALREQAAELLSVVTEHPEPLGDDGVGYSLATGRAAMEHRAVVAD 529
QY 9 -----8
Db 530 DRDSFVAGLTAAGVPAANVVGAAACKGVAFVFPQGGSHWQGMARELSSESPVFRRK 589
QY 9 -----8
Db 590 LAECAATAPIYVDWLLGLVLRGDPDAPALDRDDVTQLALFAMVSLAELWRCGVEPAV 649
QY 9 -----8
Db 650 VCHSGEITAAHVAGALSITDAVRIIARCDVAVSALTGKGMALALPESAVVKRIAGLP 709
QY 9 ELLVAV-----14
Db 710 ELTVAANGPGSTVVSSEPSALERLQELTAENVOTRRVIGIDYASHSPQIAOVQGRLLDR 769
QY 15 -----14
Db 770 LGEVSEPAEIAFYSTVTGERTDTGRLDADYWYQNLROPVROQTVMADOGYFFVEV 829
QY 15 -----14
Db 830 SPHLLTAGIQTLEAADAGGVVVGSLRRGEGSRRLWLSLAECOVRLPVNWEQVFLNT 889
QY 15 -----14
Db 890 GARRVPLPTYPFQORYWLESAEYDAGDLGVSGLLSAEHPLLGAAVTLADAGFLLTGKL 949
QY 15 -----14
Db 950 SVKTOPWLADHVVGAILLPGTAFVEMILIRAADQVGCDDLIELSLTTPVLVLPATGAVQV 1009
QY 15 -----17
Db 1010 IAVGPDENGRRSVRVHSCRODDAVPDQSWTCHATGTLTSSDHQDAGQDGIWPPNDAVA 1069
QY 18 -----VFO-----AHQEDT-----26
Db 1070 VPLDSFYARAAERGFDFGPAFOGLQAAMKRGDEIFAEVGLPTAHREDAGRFGIHPALLDA 1129
QY 27 -----26
Db 1130 ALQALGAABEDPDEGMLPFAMQVSLKATGALSRLHLPAGANAVSVPTTDTTGOAVLS 1189
QY 27 -----26

Db	1190	IDSLVLRQISDKOLAAAAAMEHESLFRVDMKRISPGAAKPFVSWAVIGNDELARACGSALG	1249
Qy	27	-----	26
Db	1250	TELHPDLTGLADPPDPPVPCASRQDLDVASEAAATQRMLDLIQDWLAAARFAGSRL	1309
Qy	27	-----ERYVLTNIN-----	35
Db	1310	VVVTGGAAGTGAEGVSDLVHAASWGLLRSAQSENDFVLVDVDTAESWRALAAAVRS	1369
Qy	36	-----	35
Db	1370	GEPOLARAGEVRVPRLARCVAAEDSRI PVPGADGTVLISGGTGLGLVARHLVAERG	1429
Qy	36	-----IGAE-----	39
Db	1430	RRVLVLAGRGWSAPGVTDLVDELVGLGAAVEVASCVDGDAQDLRLTTISAEFFLRGVV	1489
Qy	40	-----	39
Db	1490	HAAGALADGWESLTPHVAKVFGPKAAGAMHLHLLTDLDSFFVLFSSFSFGVAGAAGQ	1549
Qy	40	-----	39
Db	1550	GNVAAANAFDGLAQHRRTAGLPVAVSLAWGLWEQPSGMTGALDAAGRSRIARTNPPMSAP	1609
Qy	40	-----LLRD-----	43
Db	1610	DGLRUFEMAFRYPGESLLVPVHVLDNALRADAADCGVPALLRDLVPAPVRRSAVNESADV	1669
Qy	44	-----PSL-----CAQPR-----	51
Db	1670	NGLVGLRRLPDLDQETOLLGLVREHSAVLGSHGAVEVGADRAFRLGDFSLSGVEFRN	1729
Qy	52	-----	51
Db	1730	RLGGVLGVRLPATAVFDYPTPRALVRFLLDKLIIGVVEAPTAPAAVAATAADPPVIVGM	1789
Qy	52	-----	51
Db	1790	GCRYPCGVSSPEELWRLVAGGLDAVAEPDORGWDAQGLFDPDPRLGTSYVCEGGFLRD	1849
Qy	52	-----	51
Db	1850	AAEFDAFPFGISPREALAMPQORLLLEAVMETVERAGIDPLSLRGSRGTGVFAGLMHHDY	1909
Qy	52	-----VHL-----	54
Db	1910	GARFITRAPEGFEGLNGSAGGVFSGRVAYSFGFEGPAVTVDTCSSSLVALHLAGOAL	1969
Qy	55	-----	54
Db	1970	RSCECDLALAGGVTVMATPGMFVFSRQRLAADGRCKSFPAAADGTGCGEGAGLVLLER	2029
Qy	55	-----	54
Db	2030	LSDARRNGHAVLAVVRGSVAVNQDASNGLTAPNGPSQORVITQALASAGLSVSDVAVEA	2089
Qy	55	-----VKMV-----	58
Db	2090	HGTGTRLDGPIEAQALIAITYCGGRDSRPLWLGLSVKSNIGHTQAAAGVAGVIRKVMAMRH	2149
Qy	59	-----ILTE-----	62
Db	2150	QQLPATLHVDEPTSEVDMSAGDVOLLTENTPMPGNHPRRVGVSSFGISGCTNAHVILEQA	2209
Qy	63	-----	62
Db	2210	SKTPOETADKSGPSESTVDLPVPLIVSGRTPAALSQAASALLSYLGERGDI STLDAAF	2269
Qy	63	-----	62
Db	2270	SLASSRAALEERAVVLGADRETLTLLSGLEALASGREASGVSGSPVSGGVGFVAGOGGOW	2329
Qy	63	-----	62
Db	2330	LGMGRGLYSVPFPADAPDEACAGLD AHLGQDVGRVDRVVFSGDSGLDRTLMAOSGLFAL	2389
Qy	63	-----PEG-----	65
Db	2390	QVGLLSLLGSGVGRPGVVVLGHSVGEFAAANAAGVLSLPDAARMVAGRARLMOALPSSGAM	2449
Qy	66	-----	65
Db	2450	LAVAAGEEQLRPLADRVDGAGIAAVNAPESVWLSGDREVLDDIAGALDGGIRWRLRV	2509
Qy	66	-----	65
Db	2510	SHAFHSYRMDPMLOEFAEIAARSVDYRRGDLPVVS TLTGELDTAGMATPEYVWVQVREP	2569
Qy	66	-----	65
Db	2570	RFADGVRVLAQGVATIFELGPDATLSALIPDCHSWADQAMPI PMLRKDRTETETVVA	2629
Qy	66	-----	65
Db	2630	ARAHTRGVVIEWSAYFAGTGARRVELPTAFQORQRYWLETSDYGDVTGIGLAAAEHPLLG	2689
Qy	66	-----	65
Db	2690	AVVALADGDGMLTGRLSVGTHPMLAOHRVLGEVVVPGTATILEMALHAGARLGCORVEEL	2749
Qy	66	-----	65
Db	2750	TLETPLVVPERAGAGSRGPAGGTTVSIETAEBRVRTNDAIEIOLLVNADEGGRVRSVL	2809
Qy	66	-----	65
Db	2810	YSRPGSGRGGWTRHATGELVVGTTGGRAVPDMSAEGAESIALDEFYVALAGNGFEYGP	2869
Qy	66	-----	65
Db	2870	LFQGLQAAWRRGDEVLAIEIAPPAAEDAMASGYLLDPALLDAALQASALGDRPEQGGANLP	2929
Qy	66	-----	65
Db	2930	FSFTGVELSAPAGTISRVRLETRRRPDAISVAVMDESGRLLASIDSLRLRSSVSQLANRD	2989
Qy	66	-----	65
Db	2990	AVRDALFEVTWEPVATQSTEPGRWALLGDTACGKDDLIKLATDSADRCADLAAAEKLDS	3049
Qy	66	-----	65
Db	3050	SALVPDVVVYCAGEQADPGTGAALAEATOOTLALLOAWLAEPRLAEARLVVVTCAAVTTA	3109
Qy	66	-----	65
Db	3110	PSDCASELAHAPLWGLLRAAQVENPGQFVLADVDTAESWRALPSALGSMEPQLALRKGA	3169
Qy	66	-----	65
Db	3170	VRAPRLASVAGQIDVPVAVVADPRTVLISGGTGLLGAVARHLVTERGVRRRLVLTORRGW	3229
Qy	66	-----ANLTSLSLV-----CG-----	82
Db	3230	DAPGITELVCELNGLCVVDVVACDVADRADLESLLAAVPAEFPCLGVVHAAGALADGVI	3289
Qy	83	-----WS-----	84
Db	3290	ESLSPDDVGAVFGPKAAGANLHELTRDTDLSPFALFSSLSGVAGAPGOGNYAAANAFLD	3349
Qy	85	-----	84
Db	3350	ALAHYRRSOGLPAVSLAWGLWEOPSGMTETLSEVDRSRIRANPPILSTKEGLRFLDAGLA	3409

QY 85 ----- 84
Db 3410 LDRAAVPAKLDRTEFLAEQARSGLPALLTALVPPIRNRASGTGLADEGTLGVVREH 3469
QY 85 ----- 84
Db 3470 AAVALGYSSAADYGVVERAFDGLGDSLSGVELRNLAGVLGVRLPATAVFDYPTPRALAR 3529
QY 85 ----- 91
Db 3530 FLHQELADEIATTPAPVTTTRAPVAEDDLVAIVGMGRFPQGVSSPEELWRLVAGGVDAV 3589
QY 92 ----- 91
Db 3590 ADPPADRGWDLAGLFDPPDPERAKTYVREGAFLTDADRFDAGFFGISPREALANDPQORL 3649
QY 92 ----- 91
Db 3650 LLELSWEAIERAGIDPGSLRSGRTGVFAGLMYHDYGARFASRAPEGFEGVLGNGSAGSVA 3709
QY 92 ----- 91
Db 3710 SGRIAYSGFEGPAVTVDTACSSSLVALHLAGSLRSGECDLALAGGVTVMSPTGFVEF 3769
QY 92 ----- 91
Db 3770 SRORGLAPDGRCKSFASADGTGWGEGAGLVLLERLSDARRNGHRVLAVVRGSVAVNQDGA 3829
QY 92 ----- 94
Db 3830 SNGLTAPNGSQORVIOQALASAGLSVSDVDVAEAGTGTRGLGDPTEAQALIAITYGRDRD 3889
QY 95 PG----- 99
Db 3890 PGRPLMLGSVKSNIHTQAAAAGVAGVIKMYMAMRHGOLPRTLHVDAPESSQVDSAGRVOL 3949
QY 100 ----- 99
Db 3950 LTENTPWPDSGRPCRVGVSSFGISGTTNAHVILEQSTQGMQQAABPDSSPVLDPVVPVWV 4009
QY 100 ----- 106
Db 4010 SGKTPEALSAQAATLATYLDQNDVSDPLDVGISLAVTRSDALDERAVVLGSDRDTLLSGLN 4069
QY 107 ----- 108
Db 4070 ALAAGHEAGVWTGPVGIGGRTGVFAGQGGQWLMGMRRLYSEFPAPAGAFDEACAEIDA 4129
QY 109 ----- 108
Db 4130 NLGREVGVRDVFGSDESLDRTLMAOSGLFALQVGLWELLGTWGVRRPSVVLGHSVGELA 4189
QY 109 ----- 108
Db 4190 AFAAGVLSNAEAARLVAGRARLMQALPCSGAMLAVSATEARYGPLLDDGVRDRVGAAYN 4249
QY 109 ----- 108
Db 4250 APGSVVLGSDRDVLDGIAGRLDGGIRSRWLRVSHAPSHRMDPMLAEFAELARSVDYRS 4309
QY 109 ----- 108
Db 4310 PRLPVSTLTGNLDDGVGMATPEYWRQVRREPVRFDGVQALVDQGVDTTIVELGPDGALS 4369
QY 109 ----- 108
Db 4370 SLVQECVAESGRATGIPLVRDRDRVTVLDALAQHTRCGAVDWGSGFFAGTRATQVDLP 4429
QY 109 ----- 108
Db 4430 TYAFORQRYWLEPSDSDGTGVLGTGAEPHLLGAVVPVAGDGVLLTGRLSVGTHPWLAE 4489

QY 109 ----- 111
Db 4490 HRVLGEVVVPGTALLEMAWRAGSOVCCERVELTLEAPLVLPERGAAAQVQAVGAPDEAG 4549
QY 112 ----- 120
Db 4550 RRSQLYSRGAGEDGDWRRRIASGLLAQNAVPPADSTAMPDPGAGQVDIAEFYERLAERG 4609
QY 121 VT----- 122
Db 4610 LTYGPVFOGLRAWRHGGDDIFPAELAGSPDASGFIHPALLDAALHAMALGASPDSEARLP 4669
QY 123 ----- 125
Db 4670 FSWRGAQLVRAEGAALRVRLSPLGSGAVSLTLVDATGRRVAAVESLSTRPVSTDOIGACR 4729
QY 126 ----- 132
Db 4730 GDQERLLHVEWVRSAESAGMSLTSCAVVGLGEPEHAAKTKTGVQVESHADLASLATEVA 4789
QY 133 ----- 132
Db 4790 KRGSAPGAVIVPCPRRMOELPTAARRATQAMAMLOQWLADDRFVSTRLLILLTHRAVS 4849
QY 133 ----- 132
Db 4850 AVAGEDVLDLVHAPLWGLVRSQAEPDRFALIDMDDERASQTALAEALTAGEAQLAVRS 4909
QY 133 ----- 141
Db 4910 GVVLPRLQGVKSGEAPRWDEGTVLVTGGTGLGALLARHLVSAHGVHLLASRRGL 4969
QY 142 ----- 141
Db 4970 AAPGADELVAELQACADVAVVACDSADRSLSARLVASVPAENPLRVVTHAAAGVLDGVL 5029
QY 142 ----- 141
Db 5030 MSMSPERLDVLRPKVDAAYLWHLTRELGLSAFVLFSVAGLFGAGOSVYAGNAFLD 5089
QY 142 ----- 141
Db 5090 ALAHCRAQGLPALSLASGLWASIDGMAGDLAAADVRLSRAGIGPLSAPGGLALFDAV 5149
QY 142 ----- 141
Db 5150 GSDEPLAPVRDVEALRVQARSVQTRIPEMLHGMMAGPSRRTPTTSRVEPLHERLAGUS 5209
QY 142 ----- 141
Db 5210 EGERRQOVLQVRADIAVVLGHGRSSVDIEKPLAELGFDLSLTAIELNRNLATATGLRUP 5269
QY 142 ----- 141
Db 5270 ATLAFDHGTAAALAHQVCAQLGTATAPAPRRRTDDNDATEDPVRSLFQOAVAACRILDGMDL 5329
QY 142 ----- 141
Db 5330 VKVAAQLRPVFGSPGELESPLKPVQLSRGPEELALVCMPPALICMPPAQOYARIAGFRDV 5389
QY 142 ----- 141
Db 5390 RDVSVIPMPGFTAGEPLPSAIEAVRVTOAEAVLQEFAGGSFVLVGHSSGGWLAHEVAGEL 5449
QY 142 ----- 147
Db 5450 ERRGVVAGVWLLDVTIPGEITPRFSVMAHRTYEKLATFTDMQDVGIT 5498

RESULT 5

US-09-370-700-6

; Sequence 6, Application US/09370700

; Patent No. 6274350

```
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 5588
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-6

Query Match      27.9%; Score 216; DB 4; Length 5588;
Best Local Similarity 1.7%; Pred. No. 0.6;
Matches 92; Conservative 25; Mismatches 29; Indels 5123; Gaps 30;

Qy  2 AGGI----- 5
Db 230 AGGISVMTGPAVFESRQRALASDGRCKPFADAADGTGWGEGAGMLLLERLSDDARRNGH 289
      ||||
Qy  6 ----- 5
Db 290 PVLAAVGSAINQDTSNGLTAPSGPAQQRVIRQALANAGLSPAEDVVEAHGTCTALGD 349
Qy  6 -----LHL 8
Db 350 PIEAQAITYCANRSADHPLLGLSKNIGHTQAAAGVAGVIKSVLAIRHREMPRSLHI 409
      ||||
Qy  9 ----- 8
Db 410 DOPSHVDWSAGAVRLTDSVDWPDGRPRAGVSSFGMSGTNAHLIVEEVSDFPVSGST 469
Qy  9 ----- 8
Db 470 EPTGAPWPLSKTETALREQAAELLSVVTEHPERGLDVGYSLATGRAAMEHRAVVAD 529
Qy  9 ----- 8
Db 530 DRDSFVAGLTALAAGVPAANVVQGAADCKGVAFVFPQGGSHWQGMARELSESSPVFRK 589
Qy  9 ----- 8
Db 590 LAECAATAPYVDHSLGLVLRGDDPAPALDRDDVIQLALFAMMVSLAELWRSCGVEPAV 649
Qy  9 ----- 8
Db 650 VGHSGEIAAHVAGALSITDAVRIIAARCDASVALTCKGGMALALPESAVVKRIAGLP 709
Qy  9 ELLVAV----- 14
      ||||
Db 710 ELTVAAVNGPGSTTVSGEPSALERLQELTAENVQTRRVGIDYASHSPQIAQVQGRLLDR 769
Qy 15 ----- 14
Db 770 LGEVSEPAETAFTYSTVTGERTDTGRLDADYVYQNLROPVRFOQTVMARMADQGYRFFVEV 829
Qy 15 ----- 14
Db 830 SPHPLLTAGIQTLEAADAGGVVGSLLRRGEGSRMLTSLAECQVRCPLPVNWEQVFLNT 889
Qy 15 ----- 14
Db 890 GARRVPLPTYPFORVWLESAEYDAGDLGSLGSLSAEHPLLGAATLADAGGFLTGKL 949

Qy 15 ----- 14
Db 950 SVKTPWLAHDVVGGAILLPGTAFVEMLIRAAQVQCDLIEELSLTTPVLVLPATGAVQVQ 1009
Qy 15 -----GPD----- 17
      ||||
Db 1010 IAVGGPDEAGRRSVRVHSCRDDAVPQDSWTCHTATGTLTSSDHQDAGQDGPDIWPPNDAVA 1069
      ||||
Qy 18 -----AQEDT----- 26
      ||||
Db 1070 VPLDSFYARAABRGDFGPAFQGLQAANKRGDEIFAEVGLPTAHREDAGREGIHPALLDA 1129
      ||||
Qy 27 ----- 26
Db 1130 ALQALGAABEDPDEGMLPFAWQGVSLKATGALSRLVHLVPAGANAVSVFTTDTTGOAVLS 1189
Qy 27 ----- 26
Db 1190 IDSILVLRQISDKLAAARAMEHESLFRVDMWKRIKSPGAAPKPVSMVAVIGNDELARACGSALG 1249
Qy 27 ----- 26
Db 1250 TELHPDLTGLADPPDVVVVPCGASRQDLVASEARAATORMLDLIQDWLAAARFAGSRL 1309
Qy 27 -----ERYVLTNLN----- 35
      ||||
Db 1310 VVVTGGAATGPAEGVSDLVHAASWGLLRSAQSENPDPRFVLVDVDTGTAESWRALAAAVRS 1369
Qy 36 ----- 35
Db 1370 GEPOLALRAGEVRVPRLARCVAAEDSRIPVPGADGTVLSGGTGLLGLVARHLVAERGV 1429
Qy 36 -----ICAE----- 39
      ||||
Db 1430 RRLVAGRRGWSAPGVTDLVDELGLGAAVEASCDVGDRAQLDRLLTTISAEFPLRGV 1489
Qy 40 ----- 39
Db 1490 HAAGALADGVVESLTPEHVAKVFGPKAAGAWHLHELTLDDLDLFFVLFSSESFGVAGAAGQ 1549
Qy 40 ----- 39
Db 1550 GNYAANAFLDGLAQHRRTAGLPAVSLAWGLWEOPSGMTGALDAAGRSRIARTNPPMSAP 1609
Qy 40 -----LLRD----- 43
      ||||
Db 1610 DGLRLFEMAFRVPGESLLVPVHVDLNALRADAADGGVPALLRLDLVPAPVRRSAVNESADV 1669
Qy 44 -----PSL-----GAQPR- 51
      ||||
Db 1670 NGLVGLRRLPDLDOETQLLGLVREHVSAVLGHSGAVEGADRAFRDLGDFSLSGVEFRN 1729
Qy 52 ----- 51
Db 1730 RLGGVGLVRLPATAVFDYPTPRALVRFLDKLIGVEAPTPAPAAVAATDDPVPVVG 1789
Qy 52 ----- 51
Db 1790 GCRYPGVSPEELMWRLVAGGLDAVAEFDDRGDQAGLPDPDPDLGTSYVCEGGFLRD 1849
Qy 52 ----- 51
Db 1850 AAEFDAGFGCISPREALAMDPOORLLLEVANWETVERAGIDPLSLRGSRTGVFAGLMMHDY 1909
Qy 52 -----VHL----- 54
      ||||
Db 1910 GARFITRAPBGPFGYLGNSAGGVFSGRVAYSFGFPGVAVTVDTCSSSLVALHLAQAL 1969
Qy 55 ----- 54
Db 1970 RSGECDALAGGVTVMATPGMFVFSRORGLAADGRCKSFMAAADGTGMGEGAGLVLLER 2029
```

QY 55 ----- 54
Db 2030 LSDARRNGHAVLAVRGSVAVNODGASNGLTAPNGPSQORVITOALASAGLSVSDVAVEA 2089
QY 55 -----VKMV----- 58
Db 2090 HGTGTRLDPIEAQALATYCGORSDRPLMLGSKVNSIGHTQAAAGVAGVIMVAMR 2149
QY 59 -----ILTE----- 62
Db 2150 GOLPATLHVDEPTSEVDWSAGDVQLLTENTPMPGNHPRRVGVSSFGISGTNAHVILEQA 2209
QY 63 ----- 62
Db 2210 SKTPDETADKSGPSESTVDLPVPLIVSGRTPAALSQAQALLSYLGERGDI STLDAAF 2269
QY 63 ----- 62
Db 2270 SLASSRAALEERAVVLGADRETLISGLEALASGREASGVSGSPVSGGVGFVAGQGOW 2329
QY 63 ----- 62
Db 2330 LMGRLYSVPFVADFEACAGLDAHLGQDVGVVDFVFGSDGLDRTLWAQSLFAL 2389
QY 63 -----PEG----- 65
Db 2390 QVGLLSLLGSGVVRPGVVLGHSVGEFAAUAAGVLSLPDAARMVAGRARLMQALPSGAM 2449
QY 66 ----- 65
Db 2450 LAVAAGEEQLRPLADRVDGAGIAAVNAPESVLSGDEVLLDIAGALDQGIWRRLRV 2509
QY 66 ----- 65
Db 2510 SHAFSYRMDPMLQEFAPAIARSVDYRRGDLPVVSTLTGELDTAGVMATPEYVWRQVREP 2569
QY 66 ----- 65
Db 2570 RFADGVRVLAQGVATIFELCPDATLSALIPDCHSWADOAMPIMLRKORTETETVVAAV 2629
QY 66 ----- 65
Db 2630 ARAHTRGVPEWSAYFAGTGARRVELPTAFQRYWLETSDYGDVTGIGLAAAEHLIG 2689
QY 66 ----- 65
Db 2690 AVVALADGCMVLTGRLSVGTHPWLAQHRVLGEVVVPGTALLEMALHAGARLCCDRVEL 2749
QY 66 ----- 65
Db 2750 TLETPLVPERAAGAGSRGPAGGTVSIETAEERVRTNDAIEIQLLVNAPDEGRRRVS 2809
QY 66 ----- 65
Db 2810 YSRPAGSRGGWTRHATGELVVTGGRVAVPDWSAEGAEALDEFYVALAGNFEYGP 2869
QY 66 ----- 65
Db 2870 LFOGQAAWRRGDEVLAETAPAEADAMASGYLLDPALLDAALQASALCDRPEQGWLP 2929
QY 66 ----- 65
Db 2930 FSFTGVLSAPACTISRVLETRRPDAISVAVMDESGRLLASIDSLRSLRSVSSGQLANRD 2989
QY 66 ----- 65
Db 2990 AVRDALEVTWEPVATQSTEPGRWALLGDTACGKDDLIKLATDSADRCADLAALAEKLD 3049
QY 66 ----- 65
Db 3050 SALVPDVVYVCAGEADPGTGAALAEATQOTLALLQAWLAEPRLAEARLVVVTCAVITA 3109
QY 66 ----- 65

Db 3110 PSDGASELAHAPLWGLLRAAOVENPQOFVLADVDTAESWRALPSALGSMEPQALRKA 3169
QY 66 ----- 65
Db 3170 VRAPRLASVAGOIDVPAVVADPDRTVLISGGTGLLGGAVARHLVTERGVRRLLVLTGRROW 3229
QY 66 -APNIT-ANLTSSLSV-CG----- 82
Db 3230 DAPGITELVGLNGLVAVVDVACDVADRDLLESLLAAVPABFPLCGVVVHAAGALADGVI 3289
QY 83 -----WS----- 84
Db 3290 ELSLPPDVGAVFGPKAAGAWNLHELTRDTLSFFALFSSLSGVACAPGQNYAANAFLD 3349
QY 85 ----- 84
Db 3350 ALAHYRRSOGLPVAVSLANGLWEQPSGMTETLSEVDRSRIARANPPLSTKEGLRFLDAGLA 3409
QY 85 ----- 84
Db 3410 LDRAAVVPKLDRTFLAEQARSGSLPALLTALVPIRRNRASCTELADEGTLGCVVREH 3469
QY 85 ----- 84
Db 3470 AAALGYSSAADVGVVERAFRDLGFDLSGLSVELNRLAGVLGRLPATAVFDYPTPRALAR 3529
QY 85 -----QTINPED----- 91
Db 3530 FLHQELADEIATTPAPVTTTTPVAEDDLVAIVGMGRCPFGQVSSPBELMRLVAGGVDAV 3589
QY 92 ----- 91
Db 3590 ADFPADRGWDIAGLFDPPPERAGTKYVREGAFLTDADRFDAGFFGISPREALAMDPOORL 3649
QY 92 ----- 91
Db 3650 LLELSWEAIERAGIDPGLSGSRGTGVFAGLUMYHDYGARFASRAPEGEGYLGNGSAGVA 3709
QY 92 ----- 91
Db 3710 SGRIAYSGFEGPAVTVDTACSSSLVALHLAGOSLRSGECDLALAGGVTVMTPTGTFVEF 3769
QY 92 ----- 91
Db 3770 SRQGLAPDGRCKSFAESADGTGMEGAGLVLLERLSDARRNGHRVLAUVRGSVAVNODGA 3829
QY 92 -----DTD 94
Db 3830 SNGLTAPNGPSQORVIOQALASAGLSVSDVDAVEAHGTGTRGLGDPIDIAQMIATYGRDRD 3889
QY 95 PG-----HAD----- 99
Db 3890 PGRPLWLGSVKSNIHTQAAAAGVAVIKVMAMMRHGLPRTLHVADAPSSQVDSWAGRVOL 3949
QY 100 ----- 99
Db 3950 LTENTPMPDSGRPCRVGVSSFGISGTNAHVILEQSTQMDQAAEPDSSPVLDPVVPVWV 4009
QY 100 -----LVLYITR----- 106
Db 4010 SGTPEALSQAATLATYLDQNVDSPLDVSGISLAVTRSNALDERAVVLGSDRDTLLSGLN 4069
QY 107 -----FD----- 108
Db 4070 ALAAGHEAAGVVTGPVIGTGRTGFVAGOGQWLMGRRLYSEFPAPAGAFDEACAELOA 4129
QY 109 ----- 108
Db 4130 NLGREGVVRDVFGSDESLDRTLWAQSLFALQVGLWELLGTWGVVRPSVVLGSHVGEA 4189
QY 109 ----- 108

```

Db 4190 AAFAGVLSMAEAAELVAGRARLMOALPSGAMLAVSATSEARVGLDGVDRVGAAYN 4249
Qy 109 -----
Db 4250 APGSVVLSDGRDVLDCIAGRLDGGQIRSRWLRVSHAFSHRMDPMLAEFAELARSVDYRS 4309
Qy 109 -----
Db 4310 PRLPIVTLTGNLDDGVNATPEYVWVQVRFVADGVQALVDQGVDTTIVELGPDGALS 4369
Qy 109 -----
Db 4370 SLVQECVAESGRATGIPLVRRDRDEVRTVLDAQTHTRGGVDWGSFFAGTRATQVDLP 4429
Qy 109 -----
Db 4430 TYAFQRQRYWLEPSDGDVTGVLTAEGHPLLGAVVPVAGGDEVLLTGRLSVTHGHPWLA 4489
Qy 109 -----
Db 4490 HRLGELVVVPGTALLEMAWRAGSQVGCERVEELTLEAPLVLPERGAAVQLVAGPDEAG 4549
Qy 112 -----
Db 4550 RRSQLYSRGADECDWRRTASGLLAQANAVPPADSTAMPDPGAGQVDLAERYLAERG 4609
Qy 121 VT-----
Db 4610 LTYGVPQGLRAAHRGDDIFAELAGSPDASGFIHPALLDAALHMAWLGASPDSEARLP 4669
Qy 123 -----
Db 4670 FSWRGAQLYRAEGALVRVRLSPGSAVSLTLVDATGRRVAAVESLSTRPVSTQIGAGR 4729
Qy 126 -----
Db 4730 GDOERLLHVENVRSASAGMSLTSCAVGLGEPENHAAUKTTGVQVESHADLASLATEVA 4789
Qy 133 -----
Db 4790 KRGSAPGAVIVPCPRPMQELPTAARRATQQAAMQLQWLADDFVSTRLLILTHRAVS 4849
Qy 133 -----
Db 4850 AVAGEDVLDLVHAPLWGLVRSQAEPDRFALIDMDDERASQALAEALTAGEAQLAVRS 4909
Qy 133 -----
Db 4910 GYVLAPRLGVKVSAGEAFRNDGTVLVTGGTGGGLGALLARHLVSAHGVHLLLASRRGL 4969
Qy 142 -----
Db 4970 AAPGADELVALEQAGADVAVVCDSADRDLSLARLVASVPAENPLRVVWHAAGVLDGVL 5029
Qy 142 -----
Db 5030 MSMSPERLDVLRPKVDAANYLHETRELGLSAFVLFSSVAGLFGGAGQSYYAAGNAFLD 5089
Qy 142 -----
Db 5090 ALAHCRQAQGLPALSLASGLWASIDMGADLAAADVERLSRAGTGPLSAPGGLAFDAV 5149
Qy 142 -----
Db 5150 GSDEPLAPVLDVEALRVQARSVOTRIPEMLHGMAMPSSRRTPTFSRVEPLHERLAGLS 5209
Qy 142 -----
Db 5210 EGERRQVQVRADIAVVLGHGRSSVDVIEKPLAELGFDLSLTAIELNRNLATATGRLRP 5269
Qy 142 -----
Db 5270 ATLAFDHGTAALAHQVCAQLGTATAPAPRRRTDDNDATEPVRSLFQQAAYAGRIIDGMDL 5329

```

```

Qy 142 -----
Db 5330 VKVAAQLRPVFGSPGELESLEPKPVQLSRGPSELALVCMPLIGMPPAQQYARIAGPRDV 5389
Qy 142 -----
Db 5390 RDVSVIPMPGFIAGEPLPSAIEAVRTQAEAVLQEFAGGSFVLVGHSSGGWLAHEVAGEL 5449
Qy 142 -----
Db 5450 ERRGVVPGVAVLLDTYIPGEITPRFSVAMAHRTYEKLATFTMDQDVGIT 5498

```

```

RESULT 6
US-09-335-409-5
; Sequence 5, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goslisch, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-5

```

```

Query Match      27.0%; Score 209; DB 3; Length 7257;
Best Local Similarity 1.5%; Pred. No. 3.1;
Matches 93; Conservative 23; Mismatches 32; Indels 6165; Gaps 30;

Qy 1 AAG-----
Db 877 AGGSVSWGVPFTAGRRVPLPTYPQQRQRYWIEAPAGLGATADALAAQMFYRVDMPEM 936
Qy 5 -----
Db 937 PRSSVDSRRARSQGWVLADRGVGCEAAAAALSSQGCSCAVLHAPAEASAVAEQVQALG 996
Qy 8 -----
Db 997 GRNDQGVLYLWGLDANVEAGASAEVAKVTHLAAAPVLALIOALGTGPRSPRLWIVTRG 1056
Qy 8 -----
Db 1057 ACTVGGEPAAPCAQALMGMRVAALEHPGSGWGLVDLPDEESPTVEALVAELLSDDAE 1116
Qy 8 -----
Db 1117 DOLAFQRRRAARLARVAAPPEGNAAPVSLSAEGSYLVGTGLGALLVARWLVERGAGHL 1176
Qy 8 -----
Db 1177 VLISHHGLPDREENGROQPEVRVARIAAIEALEAQGARVTVAAVDVADAEAGMAALLAAVE 1236
Qy 16 P-----
Db 1237 PPLRGVVHAAGLLDDGLLAHODAGRLARLVLRPKVEGAWLHTLTREOPLDLFVLFSSASG 1296
Qy 24 -----
Db 1297 VFGSIGGSYAAGNAFLDALADLRRTQGLAALSIAWGLMAEGMGSGAQRREHSEASGWA 1356

```

QY 24 ----- 23
Db 1357 MPTSRALAMWLLGTRATORVVIOMDWAHAGNAPRDSRGRFWRDLVTATKEASSAVP 1416
QY 24 ----- 23
Db 1417 AVERWRNASVETRSLAYELVRGVAVGMFTDQGTLDVRRFAGQGLDSLMAVEIRKL 1476
QY 24 ----- 23
Db 1477 OGELGMPLSATLAFDPTVERLVEYLLSQALELQDRTDVRSLPATEDPIAIVGAACRF 1536
QY 24 ----- 29
Db 1537 PGVEDLESYQWLLTEGVVSTVPADRWNGADGRVPGSGEQRQTYVPRGGFLREVETP 1596
QY 30 ----- 29
Db 1597 DAAFFHISPREAMSLDPOORLLLEVSWEAIERAGODPSALRESPTGVFVGAGPNEYAERV 1656
QY 30 ----- 29
Db 1657 QELADEAAGLYSGTGNMLVAAGRSLFFLGLHGPTLAVDTACSSSLVALHLGCOSLRGE 1716
QY 30 ----- 29
Db 1717 CDQALVGVNMLSPKTFALLSRMHALSPGGRCKTFSADADGYARAEGCAVVLKRLSDA 1776
QY 30 ----- 29
Db 1777 QDRDPILAVIRGTAINHDGSSGLTVPSPQAEBALLRQALAHAGVVPADVDFVECHGTG 1836
QY 30 ----- 29
Db 1837 TALGDPPIEVRLSDVYGOARPADRPLILGAANKLHWEPAAGLAGLLKAVLALGQEQIP 1896
QY 30 ----- 29
Db 1897 AOPELGELNPLLPWEALPVAVARAAVWPRTDRPRFAGVSSFGMSGTNAHWLEEAPAVE 1956
QY 30 ----- 29
Db 1957 LWPAAPERSAELLVLSKSECALDAQAARLEHLDHPELGLGDVAFSLATTSAMNHL 2016
QY 30 ----- 29
Db 2017 AVAVTSREGLLAALSAVAQOTPGAARCTIASSRGLAFLFTGQGAQTPGMGRGLCAAW 2076
QY 30 ----- 29
Db 2077 PAPREAFDRCVALFDRELDRLREVWMAEPGSAESLLLDQTAFTQPALFTVEYALTALWR 2136
QY 30 ----- 29
Db 2137 SWGVEPELVAGHSAGELVAACVAGVFSLEDGVLVAAGRLMQGLSAGAMVSLGAPAE 2196
QY 30 ----- 29
Db 2197 VAAAVAPHAASVIAAVNGEQVVIAGVEQAVQAIAGFAARGARTKRLHVSHAGSPLM 2256
QY 30 ----- 35
Db 2257 EPMLEEFGRVAASVTYRRPSVLSVNLGKVVADLSAPGYVVRHVREAVRPADGVKALH 2316
QY 36 ----- 35
Db 2317 EAGAGTVEVGPKPTLLGLLPACLPPEPTLLASLRAGREEAAGVLEALGRLWAAGGSVS 2376
QY 36 ----- 35
Db 2377 WPGVFTAGRRVPLPTYWQORQYWPDIPEPSRRHAAADPTQGWFRVDPWEIPRSLQS 2436
QY 36 ----- 35

Db 2437 EASRGSMLVLADKGVGEAVAAALSTRGLPCVVLHAPAEATSATAELVTEAAGRSWQV 2496
QY 36 ----- 39
Db 2497 VLYLWGLDVAVVGAEASIDEIGDATRRATAPVLGLARFLSTVSCSPRLWVYTRGACIVGDE 2556
QY 40 ----- 39
Db 2557 PAIAPCOMALWGMGRVALEHFGAMGGLVDLDRASPPPOASPIDGEMLVTELLSOETEDQ 2616
QY 40 ----- 39
Db 2617 LAFRHRRHARLVAAPPOQAAPVLSAEASYLVGTGLGGLGLIVAQWLVELGARHLVL 2676
QY 40 ----- 39
Db 2677 TSRRGLPDRQANCEOQPPEIRARIAAVEALEARGARVTVAAVDVADVPEMTALVSSVBP 2736
QY 40 ----- 39
Db 2737 LRGVVHAAGVSYMRPLAETDETLLESVLURPKVAGSWLLHRLHGRPLDLFVLFSSGAAMV 2796
QY 40 ----- 39
Db 2797 GSHSOGAYAAANAFLDGLAHLRRSOSLPALSVAMGLMAEGMADAEAHARLSDIGVLPMS 2856
QY 40 ----- 39
Db 2857 TSAALSALQLVETGAAQRTVTRMDWARFAPVYTAGRRNLLSALVAGRDIAPSPPAAA 2916
QY 40 ----- 39
Db 2917 TRNWRGLSVABARVALHEI VHGAVARVLGFLDPSALDPMGFMNQGLDSLMAVEIRNLQ 2976
QY 40 ----- 39
Db 2977 AELDVLSTTLAFDHPPTVQRLVEHLLVDVLKLEDRSOTÖHVRSLASDEPIAIVGAACRFP 3036
QY 40 ----- 39
Db 3037 CGVEDLESYQWLLAEGVWSAEVPADRDAADWYDPDPEIPGRYTVTKGAFRLDLQLDA 3096
QY 40 ----- 39
Db 3097 TFFRISPREAMSLDPOORLLLEVSWEALESAGIAPDTRDSPTGVFVYAGPNEYTOTLR 3156
QY 40 ----- 39
Db 3157 GFTDGAAGLYGGTGNMLSVTAGRLSPFLGLHGPTLAMDTCSSSLVALHACOSLRLGEC 3216
QY 40 ----- 39
Db 3217 DQALVGVNVLLAPETVLLSRMRALSPOGRCKTFSADADGARGEGCAVVLKURDAQ 3276
QY 40 ----- 42
Db 3277 RAGDSILALIRGSANVHDGSPSCLTVPNGPAQOALLRQALSOAGVSPVDVDFVECHGTGT 3336
QY 43 ----- 44
Db 3337 ALGDPLEVOALSEVYGRSGDRPLVLGAANKANVAHLEAASGLASLLKAVLALRHEQIPA 3396
QY 45 ----- 44
Db 3397 QPELGELNPHLPWNTLPVAVPRKAVPWGRGRPRRAGVSFGLSGTNVHVHLEAEAPER 3456
QY 45 ----- 44
Db 3457 APAAPARVVELVLSAKSAALDAAAARLSAHLSAHPELSGLDVAFSLATTRSPMEHRLA 3516
QY 45 ----- 44

```

Db 3517 IATTSREALGALDAAAQOKTPOGAVRGKAVSSRGKLAFLFTGQGAQMPGMRGLYETWP 3576
QY 45 ----- 44
Db 3577 AFREAPDRCVALFDREIDQPLREVMMAAPGLAQAAARLDQYAYAPALFALEVALAALWRS 3636
QY 45 ----- 44
Db 3637 WGVEPHVLGHSIGELVAACVAGVFSLEDAVLVAARGRLMQALPAGGAWVAIAASEAV 3696
QY 45 -----SLGAQF----- 50
Db 3697 AASVAPHAATVSIAAVNGPDVAVIAGAEVQVLALGATFAARGIRTKRLAVSHAFHSPLMD 3756
QY 51 ----- 50
Db 3757 PMLEDQFORVAATIAAYRAPDRPVVSNVTGHVAGPEIATPEYVWRHRSVAVRFGDGAKALHA 3816
QY 51 ----- 50
Db 3817 AGAATFEVGPKPVLGLLPACLGADAVLVPSLRADRSECEVVLAAALGAWYAWGGALDW 3876
QY 51 -----RV----- 52
Db 3877 KGVFPDARRVALPMYPWORERHMDLTPRSAAAPAGIAGRWPLAGVGLCMPGAVLHVLS 3936
QY 53 -----HLV----- 55
Db 3937 IGRHPQFLGDHLVFGKVVVPGAFHVAVILSTAAERWPERATELTGVEFLKAIAEMEPQOE 3996
QY 56 ----- 55
Db 3997 VELHAVLTPEAAGDGYLFELATLAAPETERMTTHARGRVQTDGAPCALPRLEVLDBA 4056
QY 56 ----- 55
Db 4057 IQPLDFAGFLDRLSAVRIGWGLWRWLQDGRVGDSEASLATLVPTVYNAHDVAPLHPILLD 4116
QY 56 ----- 55
Db 4117 NGFAVLSLLSTRSEPDGTPPLPFAVERVWRWRAVPGRVRCGGVPRSOAFGVSFVLVDE 4176
QY 56 ----- 55
Db 4177 TGEVVAEVEGFVCRRAPREVFLQESGASTAALYRLDWPEAPLPDAPAEIRIEESWVVAA 4236
QY 56 ----- 55
Db 4237 PGSENAALATRLNRCVLAEPKGLEAALAGVSPAGVICLWEAGAHEEAPAAAQRVATEGL 4296
QY 56 ----- 55
Db 4297 SVVQALDRVRLMWVTMGAVAEAGERVQVATAPVNGLGRVTMQERPELSCTILVDLEPE 4356
QY 56 -----KWVLTPEG----- 65
Db 4357 ADAARSADVLLRELGRADDETOVAFRSGKRRVARLVKATTPGEGLLVPDAESYRLEAQKG 4416
QY 66 ----- 65
Db 4417 TLDQLRLAPQRRAPGPEVEIKVTASGLNFTVLAVLGMYPGDAGPMGGDCAGVATAVG 4476
QY 66 ----- 65
Db 4477 QGVRHVAGDAVMTLGLTHREVTVDARLVVRQAPAGLTPAQAAATVPVAFLTAWLALHDLGN 4536
QY 66 ----- 65
Db 4537 LRRGERVLIHAAAGGVGMAAVQIARWIGAEVEFATASPKWAAVQAMGVPRTHIASRTLE 4596
QY 66 ----- 65
Db 4597 FAETFRQVTGGRGVDVVLNALAGEFVDASLSLLSTGGRFLEMCKTDIRDBAAVAAAHPGV 4656

```

```

QY 66 ----- 65
Db 4657 RYRVFDILELAPDRTRILERSVVEGPAAGHLRALPVHFAITKAEAAFRFMAQARHOKV 4716
QY 66 ----- 65
Db 4717 VLLPAPSAAPLAPGTGTVLLTGGLGALGLHVARMQAQGVPHMVLTGRRGLDTPGAAKAVA 4776
QY 66 ----- 65
Db 4777 EIEALGARVTIAASDVADRNALEAVLOAIPEAMPLOQVVIHAAGALDDGVLDDEOTDRFSR 4836
QY 66 --APNIT----- 70
Db 4837 VLAPKVTGAWNLHELTAAGDLAFAVLFPSMSGLGSAGQSNYAAANTFLDALAAHRAEG 4896
QY 71 -----ANLTSS----- 77
Db 4897 LAAQSLANGPWSOGMAAGLSAALQARLARHGMGALSPAQGTALLGOALARPETQLGAMS 4956
QY 78 ----- 77
Db 4957 LDYRAASQASGAAPVPMWRALVRABARHTAAGAGOGAALARLALGALPEARRADEVKRYQAE 5016
QY 78 ----- 77
Db 5017 IARVLSMAASAVPDRPLSLDGLSLTAVELRNVLQORVGATLPATLAFDHTVDALTR 5076
QY 78 ----- 77
Db 5077 WLDDKVLVAEPVSVAKSQSPQVALDEPIAIIIGICRFFGVADPESEFWRLLEEGSDAVV 5136
QY 78 ----- 77
Db 5137 EVRHERWDIDAFYDPDPDVRGKMTTRFGGFLSDIDRFDPAPFGISPREATTMDPOORLLL 5196
QY 78 ----- 77
Db 5197 ETSWEAFERAGILPERLMGSDTGVFGLFYQYAAALAGGIEAPDGYLGTGTASVASGRI 5256
QY 78 ----- 77
Db 5257 SYVLGLKPSLTDVTACSSSLVAVHLAQALRRGECVSVALAGGVALMLTPTATFVFSRLR 5316
QY 78 ----- 77
Db 5317 GLAPDGRCKSFSAADGVGWSEGCAMLLLKPLRDAQRDGPILAVIRGTAVNQDGRSNGL 5376
QY 78 ----- 77
Db 5377 TAPNGSSQOEVIIRRALEOAGLAPADVSYVECHGTGTTLGDPIEVOALGAVLAQGRPSDRP 5436
QY 78 ----- 77
Db 5437 LVIGSVKSNIGHTQAAAGVAGVIKVALALERGLIPRSLHFDAPNPHIPWSELAVOVAAKP 5496
QY 78 ----- 77
Db 5497 VEWTRNGVPRRAGVSSFGVSGTNAHVLEEAPAAAFAPAAARSSELVLSAKSAAALDAQ 5556
QY 78 ----- 77
Db 5557 AARLSAHVVAHPPELGLDGLAFSLATTRS PMTYRLVAATSREALSAAALDTAAOQAPPA 5616
QY 78 -----LSVC-----GWS----- 84
Db 5617 ARGHASTGSAKVVVFPFGQSQMLGMGQKLLSEEPFRDALSACDRAIQAEAGWSLLAE 5676
QY 85 ----- 84
Db 5677 LADETTSQLGRIDVQPALFAIEVALSALWRSWGPEDAVVGHSMGEVAAAHVAGALS 5736

```

```
QY 85 ----- 84
Db 5737 EDVAIIICRRSLLRRISQGEMAVVELSLAEAEALLGYEDRLSVAVSNSPRSTVLAGE 5796
QY 85 ----- 84
Db 5797 PAALAEVLAILAAKGVCFCRRVKVDVASHSPQIDPLRDELLAALGELEPRQATVSMRSTVT 5856
QY 85 ----- 84
Db 5857 STIMAGPELVASYWADNVQRVPRFAEAVQSLMEDGHGLFVEMSPHILTTTSVEIRATK 5916
QY 85 ----- 84
Db 5917 REGVAVGLRRGQDERLSMLAALGALVHGOAVQWNERLFSAGGAGLRVRPLPTVPMQROR 5976
QY 85 ----- 90
Db 5977 YWVDAPTGAAGGSRFAGHAGSHPLLGEMOTLSTQRSTRVMETTLDLKRLPWLGDHRVQGA 6036
QY 91 ----- 94
Db 6037 VVFGAAYLEMALSSGAELGDPLOVSDVVLAELAFADDTFAAVQVMATEERPGRLOF 6096
QY 95 ----- 97
Db 6097 HVASRVFGHGAAPRSHARGVLRQIERAEVPARLDLAALRLARLOASAPAAATYAALAEWG 6156
QY 98 ----- 97
Db 6157 LEYGPAPQGLVELWRGEALGRVRLPEAAGSPAACRLHPALLDACPHVSSAFADRGAT 6216
QY 98 ----- 97
Db 6217 PWVPVEIGSLRWFORPSEGLWCHARSVSHGKPTDRRSTDFWVVDSTGAIVAEISGLVAQ 6276
QY 98 ----- 97
Db 6277 RLAGVRRREDDWFMEPAWPTAVPGSEVMAGRWLLIGSGGLGALHSALTAECHSVV 6336
QY 98 -----AD 99
Db 6337 HATGRGTSAAQLQALLTASFDGOAPTSVVHLGSLDERGVLDADAPDADALEESVRGCD 6396
QY 100 LVLV-----ITR----- 106
Db 6397 SVLWTVQAVAGAFRDPRLWLVTIRGAIGAAGDVSVQAAPLLGLGRVIALEHAELRCAR 6456
QY 107 ----- 106
Db 6457 IDLDPARRDGEVDELLAELLADDAEEVAFRGGERRVARLVRLPETDCREKIEPAERGP 6516
QY 107 FDLEL----- 111
Db 6517 FRLEIDSGVLLDLVLATERPPGPEVEIAVEAAGLNFLDVMRAMGIYPGDPGPVAL 6576
QY 112 ----- 111
Db 6577 GAECGRIVAMGEGVESLRIGQDVAVAPFSFCTHTVIDARMLAPRPAALTAQAALPVP 6636
QY 112 ----- 111
Db 6637 AFMTANYGLVHLGRLAGERVLIHSATGGTGLAAVQIARHLGAEIFATAGTPEKRAWLRE 6696
QY 112 -----PDGN----- 115
Db 6697 QGTAHVMDRSRSLDFAEOVLAAATKGEQDVVVLNSLSGAATDASLSLTPDGRFIELGKTDI 6756
QY 116 ----- 115
Db 6757 YADRSGLAHFRKLSYSAVDLAGLAVRRPVERVAALLAEVVDLLARGALQPLPVEIFPLS 6816
QY 116 ----- 115
```

```
Db 6817 RAADAFRMAQAQHLGKLVLALEDPDVIRVPESGVAIRADGAYLVGTGGLGGLSVAG 6876
QY 116 -----ROV--- 118
Db 6877 WLAEOGAGHLVLVGRSGAVSAEQOTAVALEAHGARVTVARADVADRAQMERILREVTA 6936
QY 119 -----RGVTQGG----- 126
Db 6937 GMPLRGVVHAAGILDGGLMQOTPARFRAVMKPVKRGALHLHALTREAPLSFFVLVASGA 6996
QY 127 -----ACPTWSCL----- 135
Db 6997 GLLGSPGOGNYAAANTFLDALAHRRACGLPALSIDWGLFADVGLAAGQONRGARLVTRG 7056
QY 136 -----ITEDTG-----PD----- 143
Db 7057 TRSLTPDEGLWALERLLDGDRTQAGVMPFDVRQWVEFFPAAASSRRLSRLMTARRVASGR 7116
QY 144 -----LGVTI 148
Db 7117 LAGDRDLLERLATAEAGARAGMLQEWRAQVSOVLRLSEGKLDVDAPLTLGLMDSLMGLE 7176
QY 144 -----LGVTI 148
Db 7177 LNRRIEAVLGITM 7189
```

```
RESULT 7
US-09-568-102-5
; Sequence 5, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-102-5
```

```
Query Match 27.0%; Score 209; DB 4; Length 7257;
Best Local Similarity 1.5%; Pred. No. 3.1;
Matches 93; Conservative 23; Mismatches 32; Indels 6165; Gaps 30;
```

```
QY 1 AAG----- 4
Db 877 AAGGSVSWPGVFPTAGRRVPLTPYPMQRQRYWIEAPAEGLGATAADALAQMRYVDWPEM 936
QY 5 -----ILH----- 7
Db 937 PRSSVDSRRARSRGWLVLADRGVGEAAAAALSSQSCCAVLTHAPAEASAVAEQVTOALG 996
QY 8 ----- 7
Db 997 GRNDWQGVLYLWGLDAVVEAGASAEVAKVYTHLAAAPVLALIOALGTGPRSPRLWIVTRG 1056
QY 8 ----- 7
Db 1057 ACTVGGEPDAPCOAALWGMGRVVALEHPGSMGGLVDLDPESPTVEALVAELSPDAE 1116
```



```
QY 8 ----- 7
Db 1117 DOLAFQGRRRARLVAAPPEGNAAPVSLSAEGSYLVGTGGLGALGLVARWLVERGAGHL 1176
QY 8 -----LE-----LLVAVG 15
Db 1177 VLISRHGLPDREWGDRDQPEVVRARIAIAEALQAQARTVAAVDVADAEGMAALLAAVE 1236
QY 16 P-----DVFOAHO----- 23
Db 1237 PPLRGVVHAAGLLDDGLLAHQDAGRLARVLRPKVEGANVLHLLTREOPLDLFLVLFSSAG 1296
QY 24 ----- 23
Db 1297 VFGSIGGSAAGNAFLDALLRRTOGLAALSIAWLAEGMGSAQORREHEASGIWA 1356
QY 24 ----- 23
Db 1357 MPTSRALAAWELLGTRATORVVIOMDWAHAGAAPRDAASGRFRWDLVTATKEASSAVP 1416
QY 24 ----- 23
Db 1417 AVERWRNASVETRYSALVELVRGVVAGVMGFTDOGTLDVVRGFAEOGLDSLMAVEIRKRL 1476
QY 24 ----- 23
Db 1477 OGELGMLSATLAFDHTVERLVEYLLSQALELQDRDTRVSVRLPATEDPIAIVGAACRF 1536
QY 24 -----EDTERY----- 29
Db 1537 PGGVEDLESYWOLLTEGVVVSTEVPADRWNGADGRVPGSGEAQRTYVPRGGFLREVET 1596
QY 30 ----- 29
Db 1597 DAAFFHISPREAMSLDPQORLLEVSWEAIERAGQDPSALRESPTGVFVGAGPNEYAERV 1656
QY 30 ----- 29
Db 1657 QELADEAGLYSGTGNMLSVAAGRLSFFGLGHGPTLAVDTACSSSLVALHLGCQLRRGE 1716
QY 30 ----- 29
Db 1717 CDQALVGGVNMLLSPKTFALLSRMHALSPGGRCKTFSADADGYARABEGCAVVVLKRLSDA 1776
QY 30 ----- 29
Db 1777 QRDRDPILAVIRGTAINHDGPSSGLTVPSGPAQEALLRQALAHAGVVPADVDFVECHGTG 1836
QY 30 ----- 29
Db 1837 TALGDPIEVRALSDVYGOARPADRPILGAANKANLGHMEPAAGLAGLLKAVLALGQEIQIP 1896
QY 30 ----- 29
Db 1897 AQPELGELNPLLPWEALPVAVARAAPVMPRTDRPRFAGVSSFGMSGTNAHVVLEEAPAVE 1956
QY 30 ----- 29
Db 1957 LWPAAPERSAELLVLSKSEGALDAQAARLREHLDMHPGLGLGDVAFSLATTRSMNHRL 2016
QY 30 ----- 29
Db 2017 AVAVTSREGLLAALSAVAQOTPPGAARCIASSSRGKLAFLFTGQGAOTPGMGRLCAAW 2076
QY 30 ----- 29
Db 2077 PAFREAFRCVALFDRELDRLPRLVWMAEPGSAESLLLDQTAFTQPALFTVEYALTALWR 2136
QY 30 ----- 29
Db 2137 SKGVEPELVAGHSAGELVAAACVGVFSLEDGVLVVAARGRLMQGLSAGGAMVSLGAPEAE 2196
QY 30 ----- 29

QY 2197 VAAAVAPHAASVSIAAVNGPEQVVIAGVEQAVQAIAAGFAARGARTKRLHVSHASHSPLM 2256
QY 30 -----VLTNLN----- 35
Db 2257 EPMLEEFGRVAASVTYRRPSVLSVNSLSGKVVADELSAPGYWVRHVRREAVRFDGCVKALH 2316
QY 36 -----:::|----- 35
Db 2317 EAGAGTFVEVGPKTLLGLLPACLPEAEPTLLASLRAGREAAAGVLEALGRLWAAGGSVS 2376
QY 36 ----- 35
Db 2377 WPGVPTAGRRVPLPTYPMQRQRYWPDIEPDSRRRAAADPTQGWFYRVDMPEIPIRSLQKS 2436
QY 36 ----- 35
Db 2437 EASRGSWLVLADKGGVGEAVAAALSTRGLPCVVLHAPETSATAELVTEAAGGRSDHQV 2496
QY 36 -----IGAE----- 39
Db 2497 VLYLWGLDVAVVGAEASIDEIGDATRRATAPVLGLARFLSTVSCSPRLWVTRGACIVUDE 2556
QY 40 -----:|----- 39
Db 2557 PAIAPCQAAALWGMGRVAALHPGAWGGLVDLDRASPQASPIDGEMLVTELLSQETEDQ 2616
QY 40 ----- 39
Db 2617 LAFRHRRHAARLVAAPPQGOAAPVLSAEASYLVGTGGLGLIIVAQWLVELGARHLVL 2676
QY 40 ----- 39
Db 2677 TSRRGLPDRQWCEQQPPEIRARIAAVEALEARGARVTVAADVADVEPMTALVSSVEPP 2736
QY 40 ----- 39
Db 2737 LRGVVHAAGVSMRPLAETDETLLSVLRPKVAGSMLLHRLHGRPLDLFVLFSSGAAVW 2796
QY 40 ----- 39
Db 2797 GSHSQGAYAAANAFDGLAHLRRSQSLPALSVANGLWAEGMADAHAHARLSDIGVLPM 2856
QY 40 ----- 39
Db 2857 TSAALSALQRLVETGAARTVTRMDWARFAPVYTARGRRNLLSALVAGRDI IAPSPAAA 2916
QY 40 ----- 39
Db 2917 TRNWRGLSVAEARVALHEIVHGAVARVLGFLDPSALDPGMPGNEOGLDSLMAVEIRNLQ 2976
QY 40 ----- 39
Db 2977 AELDVRLSTTLAFDHPFTVQRLVEHLLVDVLKLEDRSDTOHVRSLASDEPIAIVGAACRFP 3036
QY 40 ----- 39
Db 3037 GGVEDLESYWQLLAEGVVVSAEVPADRWDAADWDYDPPPEIPGRTYVTYTKGAFRLDQLRDA 3096
QY 40 ----- 39
Db 3097 TFFRISPREAMSLDPQORLLEVSWEALESAGIAPDTRDSPTGVFVGAGPNEYVYTORLR 3156
QY 40 ----- 39
Db 3157 GFTDGAAGLYGGTGNMLSVTAGRLSFFLGLHGPTLAWDTACSSSLVALHLACOSLRUGE 3216
QY 40 ----- 39
Db 3217 DQALVGVNVVLLAPETVFLLSRMRALS PDGRCKTFSADADGYARGEGCAVVVLKRLDAO 3276
QY 40 -----LUR----- 42
|||
```

Db	3277	RAGSILALIRGSVNHDPSSGLTPVNGPAQOALLRQALSQAGVSPVDVFVECHGTGT	3336
Qy	43	---DP--- 	44
Db	3337	ALGDPIEQALSEVYGPGRSDRPLVLGAAKANVAHLEAAAGSLASLLKAVLALRHEQIPA	3396
Qy	45	-----	44
Db	3397	OPELGEINPHLPWNTLPAVAVPRKAVPMGRGARPRRAGVSAGLSGNTNVHVLEAEPEV	3456
Qy	45	-----	44
Db	3457	APAARPPVELVLSAKSAAALDAAARLSAHLSSAHPELSIGDVAFLATTRSPMEHRLA	3516
Qy	45	-----	44
Db	3517	IATTSREALGALDAAAQKTPOGAVRGKAVSSRGKLAFLFTQGAQMPGMGRGLYETWP	3576
Qy	45	-----	44
Db	3577	AFREAFRCVALFDREIDQPLREVMWAAPOGLAQARLDQYAYAQPALFALEYALALWRS	3636
Qy	45	-----	44
Db	3637	WGEVPHVLLGHSIGELVAACVAGVFSLEDAVRLVAARGRLMQALPAGGAMVAIAASEAEV	3696
Qy	45	-----SLGAQF----- : : :	50
Db	3697	AASVAPHAATVSIAAVNGPDVAVIAGAEVQVLALGATFAARGITRKLAESHAFHSPLMD	3756
Qy	51	-----	50
Db	3757	PMLDFORVAATTAYRAPDRPVVSNVTGHVAGPEIATPEYVWRHRSVAVRFGGAKALHA	3816
Qy	51	-----	50
Db	3817	AGAATFVEVGPKVLLGLLPACLGEADAVLVPSLRADRSCEVVLAAALGAWYAWGGALDW	3876
Qy	51	-----RV----- 	52
Db	3877	KGVPDGGARRVALPMYQWERHMDLTPRSAAPAGIAGRWPAGVGLCPGAVLHHVLS	3936
Qy	53	-----HLV----- 	55
Db	3937	IGPRHQPFGLDHLVFGKVVVPGAFHVAVILSIAAERPRAIELTGVFELKATAMEPDOE	3996
Qy	56	-----	55
Db	3997	VELHAVLTPEAAGDGYLFELATLAAPEPERRTTTHARGVQPTDGPAGALPRLEVLDDRA	4056
Qy	56	-----	55
Db	4057	IQPLDFAGFLDRLSAVRIGMWPLMRMLQDGRVGEASLATLVPTYRNAHDVAPLHPILLD	4116
Qy	56	-----	55
Db	4117	NGFAVLSLSTRSEDDGTPLPFAVERVWRAPVCRVRCGGVPRSQAFGVSSFLVDE	4176
Qy	56	-----	55
Db	4177	TGEVVAEVEGVCCRAPPREVFLRQESGASTAALYRLDWPEAPLPDAPABRIEBSWVVA	4236
Qy	56	-----	55
Db	4237	PGSEMAAALATRLNRCVLAEKPLEAALAGVSPAGVICLWEAGAHEEAPAAQVRATEGL	4296
Qy	56	-----	55
Db	4297	SVVQALDRRAVRLWVMTGAVAVEAGERVQVATAPVNWGLGRVTWQERPELSCTLVDLEFE	4356
Qy	56	-----KMVILTEPEG----- : : :	65
Db	4357	ADAARSADVLLRELGRADDETQVAFRSGKRRVARLVKATTPTEGLLVPDAESYRLEAGQK	4416
Qy	66	-----	65
Db	4417	TLDQLRLAPQORRAPGPEVEIKVTASGLNFRFTVLAVLCMYPGDAGPMGGCAGVATAVG	4476
Qy	66	-----	65
Db	4477	QQRHVAVGDAVMTLGTTLHRFVTVDARLVVRQOPAGLTTPAQAAATVPVAFLTAWLALHDLGN	4536
Qy	66	-----	65
Db	4537	LRRGERVLIHAAAGVGMAAQIARWIGAEVPATASPSKWAAVQAMGVPRTHIASRRTLE	4596
Qy	66	-----	65
Db	4597	FAETFRQVTGGRGVDVNLNALAGEFVDASLSLLSTGGRFLEMGTDIRDRAVAAAHPGV	4656
Qy	66	-----	65
Db	4657	RYRVFDILELADPDRTREILERVVVEGFAAGHLRALPVAFAITKABEAFRPMQAARHOGKV	4716
Qy	66	-----	65
Db	4717	VLLPAPSAAPLAPTGTVLLTGGGLGALGLHVARWLAQQGVPHMVLTRRGRLDTPGAAKAVA	4776
Qy	66	-----	65
Db	4777	EIEALGARVTIAASDVADRNALAVLOAIPAEMWPLQGVIIHAAGALDDGVLDQTTDRFSR	4836
Qy	66	---APNIT--- :	70
Db	4837	VLAPKVTGANLHELHETAGNDLAFVLFSSMSGLLGSAGOSNYAAANTFLDALAAHRAEG	4896
Qy	71	-----ANLTSSL----- : :	77
Db	4897	LAQSLAWGPWSGCGMAAGLSAALQARLARHGMGALSPOGTALLGQALARPETQLGAMS	4956
Qy	78	-----	77
Db	4957	LDVRAASOASGAAPVPMWRALVRABARHTAAGAGALAAARLGALEARRADEVRKVQAE	5016
Qy	78	-----	77
Db	5017	TARVLSWASASAVPVDRPLSDGLDLSLTAVELRNVLQORVGATLPATLAFDHTPTVDALTR	5076
Qy	78	-----	77
Db	5077	WLLDKVLAVABPSVSSAKSSPOVALDEPIAIIIGICRPPGVADPESFWRLLLEGSDAVV	5136
Qy	78	-----	77
Db	5137	EVPHERWDIDAFYDPPDPVRGKMTTRFGGFLSDIDRFDPAPFGISPREATTWDPQOORLLL	5196
Qy	78	-----	77
Db	5197	ETSWEAFERAGILPERLMGSDTGVFVGLFYOEYAAALAGGIEAFDGYLGTGTGTASVASGR	5256
Qy	78	-----	77
Db	5257	SYVGLKGPSLTVDTACSSSLVAVHLAQALRRGECVALAGGVALMLTPATFVFSRLR	5316
Qy	78	-----	77
Db	5317	GLAPGRCKSFSAADGVGWSEGCAMLLLKPLRDAQRDGPILAVIRCTAVNOGRSNGL	5376
Qy	78	-----	77
Db	5377	TAPNGSSQOEVIIRRALEOAGLAPADVSVYECGTGTTLLGDPLEVOALCAVLACQGRPSDRP	5436
Qy	78	-----	77
Db	5437	LVIGSVKSNIGHTQAAAGVAGVIKVALALERGLIPRSLHFDAPNPHIPWSELAVQAAKP	5496

```

QY 78 ----- 77
Db 5497 VEWTRNGVPRRAGVSSFGVSGTNAHVLEEAPAAAFAPAAARSAEFLVLSAKSAALDAQ 5556
QY 78 ----- 77
Db 5557 AARLSAHVVAHPGLGLGLAFSLATTSRPMYRLAVALTSREALSALDAAQOQAPPA 5616
QY 78 -----GWS----- 84
Db 5617 ARGHASTGSAKVVVFPQGSQWLMGQKLLSEEPFRDALSACDRAIOAEGWLLAE 5676
QY 85 ----- 84
Db 5677 LAADETTSQGRIDVQPALFAIEVALSALWRSWGPDAVVGHSMEVAHAHVACALSL 5736
QY 85 ----- 84
Db 5737 EDVAIIICRRSLLRRISGOGEMAVVLSAEAEALLGYEDRLSVAVSNSPRSTVLAGE 5796
QY 85 ----- 84
Db 5797 PAALAEVLAIALAAGVFCRRVKVDVASHSPQIDPLRDELLAALGELEPRQATVSMRSTVT 5856
QY 85 ----- 84
Db 5857 STIMAGPELVASYWADNVQRPVFAEAVOSLMEDGHGLFVEMSPHILTTTSVEEIRATK 5916
QY 85 ----- 84
Db 5917 REGVAVGSLRRGODERLSMLEALCALWVHQAVGWERLFSAGAGLRVPLPTYPQORER 5976
QY 85 -----QTINPE----- 90
Db 5977 YWVDAPTGAAGGSRFAHAGSHPLLGMOTLSTQSTRVWETTLDKRLPWLGDHRVQGA 6036
QY 91 -----DDTD----- 94
Db 6037 VVFPGAAYLEMALSGAEALGDGQLQVSDVLAELAFADDTTAAVQVMTAETERPGRLOF 6096
QY 95 -----PGH----- 97
Db 6097 HVASRVPGHGAARSHARGVLRQIERAEVPARLDLAALRARLQASAPAAATYAALABMG 6156
QY 98 ----- 97
Db 6157 LEYGPAGLVELNRGEALGRVRLPEAAGSPAACRLHPALLDACFHVSSAFADRGBAT 6216
QY 98 ----- 97
Db 6217 PHVPEIGSLRWFQPSGELMCHARSVSHGKPTPDRRSTDFWVVDSTGAIVAEISGLVAQ 6276
QY 98 ----- 97
Db 6277 RLAGVRRREDDNFMPEWPEPTAVPGSEVMAGRWLLIGSGGLGAALHSALTAEAGSVV 6336
QY 98 -----AD 99
Db 6337 HATGRGTSAGLQALLTASFQCGAPTSVVHLGSLDERGVLDADAPFDADALEERSLVRGCD 6396
QY 100 LVLY-----ITR----- 106
Db 6397 SVLMTVQAVAGAFRDPPLRLVTRGAQIGAGDVSVQAAPLLGLGRVIALEHAELRCAR 6456
QY 107 ----- 106
Db 6457 IDLPARRDGEVDELLAELLADDAEEVAFRGGERVARLVRLPETDCREKIEPAGRP 6516
QY 107 FDLEL----- 111
Db 6517 FRLEIDGSGVLLDLVLRATERPPGCEVEIAVEAAGLNFLDVMRWANGIYPPGPGDGPVAL 6576
QY 112 ----- 111

```

```

Db 6577 GAECGRIVAMGEGVESLRIGQDVVAVAPPSFGTHVTIDARMLAPRPAALTAQAALPV 6636
QY 112 ----- 111
Db 6637 AFMTAWYGLVHLGRRLRAGERVLIHSATGGTGLAAVQIARHLGAEIFATAGTPKEKRWLRE 6696
QY 112 -----PDGN----- 115
Db 6697 OGIAHVMDSRSLDFAEQVLAATKGEVDVVLNSLGAASLSTLVPDGRFIELGKTDI 6756
QY 116 ----- 115
Db 6757 YADRSGLAHFRKSLSYSVADLAGLAVRRPERRVAALLAEVVDLLARGALQPLPVEIFPLS 6816
QY 116 ----- 115
Db 6817 RAADAFRMAOAHGLKGLVLALEDPDVRIIRVPGESGVAIRADGAYLVTTGGLGLGLSVAG 6876
QY 116 -----ROV----- 118
Db 6877 WLABOGAGHLVLRSGAVSAEQQTAVAALEAHGARVTVARADVADRAQMERILREVTS 6936
QY 119 ---RGVTQLGG--- 126
Db 6937 GMPLRGVVHAAGILDDGLLMOQTTPARFRAVMKPVKGALHLHALTREAPLSPFVLYASGA 6996
QY 127 ---ACSPWMSCL--- 135
Db 6997 GLLGSPQGNYYAANTFLDALAHHRAOGLPALSIDMGLFADVLGAAGQONRGARLVTRG 7056
QY 136 ---ITEDTG---PD----- 143
Db 7057 TRSLTDEGLWALERLLDGDRTQAGVMPDFVQWVEFYPAASRRRLSRLMTARRVASGR 7116
QY 144 ----- 143
Db 7117 LAGDRLLERLATAEAGABAGMLQEVVRAQVSOVLRLSEGKLDVDAPLTSGLMDSLMGL 7176
QY 144 -----LQVTI 148
Db 7177 LRNRIEAVLIGTM 7189

```

```

RESULT 8
US-09-567-969-5
; Sequence 5, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goeblach, Joern
; FILE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIOR APPLICATION NUMBER: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-969-5

```

```

Query Match 27.0%; Score 209; DB 4; Length 7257;
Best Local Similarity 1.5%; Pred.No.3.1;
Matches 99; Conservative 23; Mismatches 32; Indels 6165; Gaps 30;

```

```
QY 1 AAGG-----4
Db 877 AAGGSVGPVFTAGRRVFLPTYWQRYWIEAPAEGLGATAADALAQMFYRVDMPEM 936
QY 5 -----ILH-----7
Db 937 PRSSVDSRRARRSGMWLADRGVGEAAAAAALSSQGSQSCAVLHAPAEASAVAEQVTOALG 996
QY 8 -----7
Db 997 GRNDQGVLYLWGLDAYVEAGASAEVAKVTHLAAAPVLALIOALGTGPRSPRLWIWTRG 1056
QY 8 -----7
Db 1057 ACTVGGEPDAAPCOAALWGMGRVAALEHPGSGWGLVDLDPESPTVEVEALVAELLSFDAE 1116
QY 8 -----7
Db 1117 DOLAFROGRRRAARLVAAPPEGNAAPVLSAEGSYLVGTGGLGALGLLVARWLVERCAGHL 1176
QY 8 -----LE-----LLVAVG 15
Db 1177 VLISRHLDPREWRGDPPEVRARIAAIEALEAQGARVTVAADVADAEGBMAALLAAVE 1236
QY 16 P-----DVFOAHO-----23
Db 1237 PPLRGVVHAAGLLDDGLLAHQDAGRLARLVRPKVEGANVWLHTLTREOPLDLFLVFSSASG 1296
QY 24 -----23
Db 1297 VFGSIGGSYAAGNAFLDALDURRTQGLAALSIAWGLMAEGMGSGAQORREHEASGIWA 1356
QY 24 -----23
Db 1357 MPTSRALAAMENLLGTRATORVVIOMDWAHAGAAPRDASGRFMDRLVTATKEASSAVP 1416
QY 24 -----23
Db 1417 AVERWRNASVETRSALYELVRGVVAGVMGFTDQGLDVRRGFABEQGLDSLMAVEIRKEL 1476
QY 24 -----23
Db 1477 QGELGMPLSATLAFDHTPTVERLVEYLLSQALELQDRDTRSVRLPATEDPIAVGAACRF 1536
QY 24 -----EDTERV-----29
Db 1537 PGVEDLESYQLLTSGVVVSTEVPADRWNGADGRVPGSGEAQRTYVPRGGFLREVETF 1596
QY 30 -----29
Db 1597 DAAPPHISPREAMSLDPQORLLLEVSWEAIERAGODPSALRESPTGCVFVGAGPNEYAERV 1656
QY 30 -----29
Db 1657 QELADEAAGLYSGTGNMLSYAAGRLSFFLGLHPTLAVDTACSSSLVALHLCQSLRRGE 1716
QY 30 -----29
Db 1717 CDQALYGVGNMLLSPKTFALLSRMHALSFGGRCKTFSADADGYARAEBCAVVVLKRLSDA 1776
QY 30 -----29
Db 1777 QORDRDPILAVIRGTAINHDCPSSGLTVPSGPAQEALLRQALAHAGVVPADVDVFECHGTG 1836
QY 30 -----29
Db 1837 TALGDPFIEVRALSDVYGOAPADRPRLILGAAKANLGHMEPAAGLAGLLKXIALGQEQIP 1896
QY 30 -----29
Db 1897 AQPELGEPLNLLPWEALPVAVARAAPVWPRTDRPRFAGVSSFGMSGTNAHVVLBEAPAVE 1956
QY 30 -----29

Db 1957 LWPAPERSABELLVLSGKSEGALDAQARLEHLDHMIPELGLDVAFSLATTRSAMNHL 2016
QY 30 -----29
Db 2017 AVAVTSREGLLAALSAVAQOTPPGAARCIASSSRGKLAFFLTQGAOTPGMGRGLCAAM 2076
QY 30 -----29
Db 2077 PAFREAFDRCVALFDRELDRLPREVMMAEPGSAESLLLDOTAFQTQPALFTVEYALTALWR 2136
QY 30 -----29
Db 2137 SWGVEPELVAGHSAGELVAACVAGVFSLEDCVRLVVAARGRLMOGLSAGGAMVSLGAPEAE 2196
QY 30 -----29
Db 2197 VAAAVAPHAASVSIAAVNGPEQVVIAGVEQAVQAIAGFAARGARTKRLHVSHASHSPLM 2256
QY 30 -----VLTNLN-----35
Db 2257 EPMLEEFGRVAASVTYRRPSVSLVSNLSGKVADELSPGYVVRHVREAVRFADGVKALH 2316
QY 36 -----35
Db 2317 EAGAGTFVEVGPKPTLLGLLPACLPEAPEPTLLASLRAGREAAAGVLEALGRLWAAGGSVS 2376
QY 36 -----35
Db 2377 WPGVFPTAGRRVPLPTYWQRYWPDIEPDSRRHAAADPTQGFYRVDMPEIPRSLOKS 2436
QY 36 -----35
Db 2437 EASRGSWMLVLADKGGVGEAAALSTRGLPCVVLHAPAEATSATLAEALVTEAAGGRSDQV 2496
QY 36 -----IGAE-----39
Db 2497 VLYLWGLDVAVVGABASIDEIGDATRRATAPVLGLARFLSTVSCSPRLVVVTRGACIVGDE 2556
QY 40 -----39
Db 2557 PAIAPCOALWGMGRVAALEHPGAWGLVDLDPASPPQASPIDGEMLVTELLSQETEDQ 2616
QY 40 -----39
Db 2617 LAFRHRRRHAARLVAAPPOGQAAAPVLSAEASYLVGTGGLGLIIVAQWLVELGARHLVL 2676
QY 40 -----39
Db 2677 TSRRGLPDRQAWCEQOPPEIRARIAAVEALEARGARTVVAADVADVEPNAALVSSVEPP 2736
QY 40 -----39
Db 2737 LRGVVHAAGVSMRPLAETDETLLSVLRPKVAGSWLLHRLHGRPLDLFLVLFSSGAAYV 2796
QY 40 -----39
Db 2797 GSHSQGAYAAANAFDGLAHLRRSQSLPALSVANGLWABEGMADAENHARLSDIGVLPMS 2856
QY 40 -----39
Db 2857 TSAALSALQRLVETGAARTVTRMDWARPAPVYTARGRRNLLSALVAGRDIAPSPPPAAA 2916
QY 40 -----39
Db 2917 TRNWRGLSVAEARVALHEI1VHGAVARVLGFLDPSALDPGMGFNEQGLDSLUMAVEIRNLLQ 2976
QY 40 -----39
Db 2977 AELDVRLSTTLAFDHPVTQRLVEHLLVDVLKLEDRSDTOHVRSLASDEPIAIVGAACRFP 3036
QY 40 -----39
```

```
Db 3037 GGVEDLESYWQLLAEGCVVVAEVPADRWDAADWDYDPDEIPGRTVYTKGAFLRDLQLDA 3096
Qy 40 ----- 39
Db 3097 TFFRISPREAMSLDPQORLLLEVSWEALESAGIAPDTLRSPTGTVFGAGNYYTORLR 3156
Qy 40 ----- 39
Db 3157 GFTDGAAGLYGGTGMLSVTAGRLSFFLGLHGPTLAMDTACSSSLVALHLACOSLRIGEC 3216
Qy 40 ----- 39
Db 3217 DOALVGGVNVLLAPETFVLLSRMRALSPDGRCKTFSADADGYARGEGCAVVVLKRLRDAQ 3276
Qy 40 ----- 42
Db 3277 RAGDSIILALIRGSANVHDGSSGLTVPGPAQQLLRQALSQAGSPVDVFVECHGTGT 3336
Qy 43 --Dp----- 44
Db 3337 ALGDPIEQALSEVYGPGRSGDRPLVLGAAKANVAHLEAASGLASLLKAVLALRHEQIPA 3396
Qy 45 ----- 44
Db 3397 QPELGELNPHLPWNLTLPVAVPRKAVPNWRCARPRRAGVSAGFSLGNTNVHVLEEAPVEP 3456
Qy 45 ----- 44
Db 3457 APAAPARPVELVLSAKSAAALDAAAARLSAHLASHPELSGDVAFSLATTRSPMEHRLA 3516
Qy 45 ----- 44
Db 3517 IATTSREALRGALDAAAQOKTPOGAVRGKAVSSRGKLAFLFTGGAQMPGMGRGLYETWP 3576
Qy 45 ----- 44
Db 3577 AFREAFDRCVALFDREIDQPLREVWMAAPGLAQAAARLDQTAQAPALFALEYALAAWLRS 3636
Qy 45 ----- 44
Db 3637 WGVPEPHVLHGHSIGELVAACVAGVFSLEDAVRLVAARGRLMOALPAGGAMVAIAASEAEV 3696
Qy 45 ----- 50
Db 3697 AASVAPHAATVSIAAVNGPDVAVIAGAEVQVLALGATTAARGIRTKRLAVSHAFHSPJMD 3756
Qy 51 ----- 50
Db 3757 PMLEDFORVAATIAYRAPDRPVVSNVTGHVAGPEIATPEYVWVRHRSVAVRFGDCAKALHA 3816
Qy 51 ----- 50
Db 3817 AGAATFVEVGPKPVLGLLPACLGEADAVLVPSLRADRSECEVVLAALGAWYAWGGALDW 3876
Qy 51 ----- 52
Db 3877 KGVFPDGAARRVALPMYPMQORRHWMDLTPRSAAAPAGIAGRWPAGVGLCNPAGVHLHVLS 3936
Qy 53 ----- 55
Db 3937 IGPRHQPLGDHLVFGKVVVPGAPHVAVILSIAAERWPERAIETGTVEFLKAIAMEPDQE 3996
Qy 56 ----- 55
Db 3997 VELHAVLTPEAAGDGYLPFELATLAA PETERRWTHARGRVQPTDGAPGALPRLEVLEDRA 4056
Qy 56 ----- 55
Db 4057 IQPLDPAGFLDRLSAVRIGMWGLWRWLQDGRVGDSEASLATLVPTYNAHDVAPLHPILLD 4116
Qy 56 ----- 55
Db 4117 NGFAVLLSTRSEPDGDTPLPFAVERVWRWAPVGRVRCGVPGRSQAGVSSFFLVDE 4176
Qy 56 ----- 55
Db 4177 TGEVVAEVEGFCRRAPREVFLROESGASTAALYRLDWEAPLDPAPAEITEESWVVVVA 4236
Qy 56 ----- 55
Db 4237 PGSEMAAALATLRLNRCVLAEPKLEAALAGVSPAGVICLWEAGAHEEAPAAQORVATEGL 4296
Qy 56 ----- 55
Db 4297 SVVQALDRDRAVRLWVMTMGAVAVBAGERVQVATAPVNLGRTVMQERPELSCTLVLDLEPE 4356
Qy 56 ----- 65
Db 4357 ADAARSADVLLRELGRADDETQVAFRSGKRVARLVKATTPGGLLVDAESYRLAQCKG 4416
Qy 66 ----- 65
Db 4417 TLDQLRLAPAQORRAPGCEVEIKVTASGLNFRTVLAVLGMYPGDAGPMGDCAGVATAVG 4476
Qy 66 ----- 65
Db 4477 QGVRHVAVGDVMTLGTLHRFVTVDARLVVRQPAGLTPAQAAVTPVPAFLTAMIALHDLGN 4536
Qy 66 ----- 65
Db 4537 LRRGERVLIHAAAGGVMAAVQIARWIGAEVFATASPKWAAVQAMGVPRTHIASRTILE 4596
Qy 66 ----- 65
Db 4597 FAETFRQVTGRCGVVNLNLAGFVDASLSLLSTGGRFLEMKGKTDIRDRAAAVAAHPGV 4656
Qy 66 ----- 65
Db 4657 RYRVFOILELAPDRTRTEILERVVEGFAAGHLRALPVHAFITKAEAAFRFMAQARHOGKV 4716
Qy 66 ----- 65
Db 4717 VLLPAPSAAPLAPTGTVLLTGGALGALHVARWLAQQGVPHMVLTRRRGLDTPGAAKAVA 4776
Qy 66 ----- 65
Db 4777 EIEALGARVTIAASDVADRNALEAVLQAI PAEMPLQGVIIHAAGALDDGVLDDEOTTDRFSR 4836
Qy 66 --APNIT----- 70
Db 4837 VLPKVTGAWNLHELTAGNDLAFVLFSSMSGLLSAGQSNYAAANTFLDALAAHRAEG 4896
Qy 71 -----ANLTSSL----- 77
Db 4897 LAAQSLANGPWSDGCGMAAGLSAALQARLARHGMGALS PAOCTALLGOALARPETOLGAMS 4956
Qy 78 ----- 77
Db 4957 LDVRAASQASGAAPPVWVRALVRAEARHTAAGAQAALALGALPEARRADEVRKVQAE 5016
Qy 78 ----- 77
Db 5017 IARVLSWASAASAVPVDRLPLSDLGLDLSUTAVELRNLGQVRGATLPATLAFDHTPTVDALTR 5076
Qy 78 ----- 77
Db 5077 WLDDKVLVAEPPSVSSAKSPQVALDEPIAIIIGICRFPQGVADPESFWLLLEGGSDAVV 5136
Qy 78 ----- 77
Db 5137 EVPHERMDIOAFYDDPDVVRGKMTTRFGGFLSDIDRFPAPFGISPREATTMPQORLLU 5196
Qy 78 ----- 77
Db 5197 ETSWEAFERAGILPERLMGSDTGVFVGLFYQBYAALAGGIEAFDGYLGTGTTASVASGRI 5256
```

QY 78 ----- 77
Db 5257 SYVLGKPSLTVDTCSSSLVAVHLACALRRGECVALAGVALMLTPATFVFSRLR 5316
QY 78 ----- 77
Db 5317 GLAPDGRCKSFAADGVGSECCAMLLKPLRDAORDGDPILAVIRGTAVNQDGRSNGL 5376
QY 78 ----- 77
Db 5377 TAPNGSSQOEIVIRRALEQAGLAPADVSYVECHGTGTLGDPTEVOALGAVLAQGRPSDRP 5436
QY 78 ----- 77
Db 5437 LVIGSVKSNIGHTQAAAGVAGVQKVALALERGLIPRSULHFDAPNPHIPWSELAVQVAAPK 5496
QY 78 ----- 77
Db 5497 VENTRNGVPRRAGVSSFGVSGTNAHVWLEEPAAAFAPAAARSAEFLVLSAKSAALDAQ 5556
QY 78 ----- 77
Db 5557 AARLSAHVVAHPGLGLDLAFSLATRSPTMYRLAVAATSREALSAAALDTAAOQAPPAA 5616
QY 78 ----- 77
Db 5617 ARGHASTGAPKVFVFPQGSQWLGMCQKLLSEBPVPRDALSACDRAIOABAGWSLLAE 5676
QY 85 ----- 84
Db 5677 LAADETTSQLRIDVVPALFAIEVALSALWRSWGPEDPAVVGHSMEVAAAHVAGALS 5736
QY 85 ----- 84
Db 5737 EDAVAICRRSILLRRISQGEAMAVVELSLAEAEALLGYEDRLSVAVSNSRSTVLAGE 5796
QY 85 ----- 84
Db 5797 PAALAEVLAILAAKGVFCRRVKVDVASHSPQIDPLRDELLAALGELEPRQATVSMRSTVT 5856
QY 85 ----- 84
Db 5857 STIMAGPELVASYWADNVQVRFPAEAVQSLMEDGHGLFVEMSPHPILTTSVEEIRATK 5916
QY 85 ----- 84
Db 5917 REGVAVGLRRGODERLSMLLEALCALWVHQAAGVWBERLFSAGGAGLRVPLTPYMQRER 5976
QY 85 ----- 90
Db 5977 YWVDAPTGGAGGSRFNAHAGSHPLLGEMOTLSTQSRSTRVWETTLDLKRLPWLGDHRVQA 6036
QY 91 ----- 94
Db 6037 VVFPGAAYLEMALSSGAELGDPLOVSDVVLAEALAFADDTPAAVQVQWATEERPGRLQF 6096
QY 95 ----- 97
Db 6097 HVASRVPGHGGAARSHARGVLRQIERAEVPAFLDLAALRLARLQASAPAAATYAAAEWG 6156
QY 98 ----- 97
Db 6157 LEYGAFOGLVELWRGEALGRVRLPEAAGSPAACRLHPALLDACFHVSAFADRGAT 6216
QY 98 ----- 97
Db 6217 PWVPEIGSLRFORPSGELWCHARSVSHGKPTDPRSTDFWVVDSTCAIVAEISGLVAQ 6276
QY 98 ----- 97
Db 6277 RLAGGVRRREDDWFPEPAWEPTAVPGSEVMAGRWLLIGSGGLGAALHSALTBAHSV 6336
QY 98 ----- AD 99

Db 6337 HATGRGTSAGLQALLTASFQGAQPTSVVHLGSLDERGVLDADAPFDADALEESLVRGCD 6396
QY 100 LVLY-----ITR----- 106
Db 6397 SVLWTVQAVAGAFRDPRLMLVTRGAQAIGAGDVSAQAPLLGLGRVIALEHAELRCAR 6456
QY 107 ----- 106
Db 6457 IDLDPARRGEVDELLAELLADDABEEVAFRGERRVARLVRRLPETDCREKIEPAER 6516
QY 107 FOLE----- 111
Db 6517 FRLEIDGSGVLDLVLRAERRPPGPGVEIATAENAGLFLDVMRAMGIYPGPGDGPVAL 6576
QY 112 ----- 111
Db 6577 GAECGRIVAMGEGVESURIGQVVAVAPFSFGTHVTIDARMLAPRPAALTAQAALPV 6636
QY 112 ----- 111
Db 6637 AFMTAWYGLVHLGRRLRAGERVLIHSATGGTGLAAVOIARHLGAEIFATAGTPEKRAWLRE 6696
QY 112 -----PDGN----- 115
Db 6697 OGIAHVMDSRSLDFAEQVLAATKGEVDVVLNSLGAADASLSLTLVPDGRFTELKGTDI 6756
QY 116 ----- 115
Db 6757 YADRSGLAHPKRSLSYSAVDLAGLAVRRPERRVAALLAEVVDLLARGALQPLPVEIPPLS 6816
QY 116 ----- 115
Db 6817 RAADAFKMAQAQHLGKLVLALEDPDVIRVPGESGVAIADGAYLVLTGGLGGLSVAG 6876
QY 116 -----ROV----- 118
Db 6877 WLABQAGHLVLVGRSAGVSAEQOTAVAALAEHAGARTVARADVADRAQMERITREVTAS 6936
QY 119 -----RGVTOLGG----- 126
Db 6937 GMPRLGVVHAAGLDDGLLWQOTPARFRAVMAKPVKRGALHLHALTREAPLSFFVLVYASGA 6996
QY 127 -----ACSPWMSCL----- 135
Db 6997 GLLGSPGQGYAAANTFLDALAHRRAOGLPALSIDWGLFADVGLAAGQONRGARLVTRG 7056
QY 136 -----ITEDTG-----FD----- 143
Db 7057 TRSLTPDEGLWALERLLDGRTOAGVMPDFVQWVEFYPAASRRSLRMTARRVASGR 7116
QY 144 ----- 143
Db 7117 LAGDRLLERLATAEAGARAGMLQEVVRAQVSOVLRLSEGKLDVDPALPLSLGMDSLMGL 7176
QY 144 -----LGVTI 148
Db 7177 LNRNREAVLGITW 7189

RESULT 9
US-09-568-480-5
; Sequence 5, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A

```

; CURRENT APPLICATION NUMBER: US/09/568,480
; CURRENT FILING DATE: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-5

```

```

Query Match      27.0%; Score 209; DB 4; Length 7257;
Best Local Similarity 1.5%; Pred. No. 3.1;
Matches 93; Conservative 23; Mismatches 32; Indels 6165; Gaps 30;

QY 1 AAGG----- 4
    |||
Db 877 AAGGSWPGVFTAGRRVPLTPYQORQYWEAPAEGLGATADALAQWFYRVDPWPM 936
QY 5 -----ILH----- 7
    |||
Db 937 PRSSVDSRRARRSGWLADRGVGEAAALSSQGCSCAVLHAPAEASAVAEQVTOALG 996
QY 8 ----- 7
Db 997 GRNDWQGVLYLWGLDAVVEAGASAEVAKVTHLAAAPVLIALQALGTGPRSLWVTRG 1056
QY 8 ----- 7
Db 1057 ACTVGGEPAAPCOAALMGMRVALEHFGSGWGLVDLPDESPTEALVAELLSPOAE 1116
QY 8 ----- 7
Db 1117 DQAFRQGRRAARLVAAPPEGNAAPVLSAEGSVLTGGLGALLVARWLVERGAGHL 1176
QY 8 -----LLVAVG 15
    |||
Db 1177 VLISRHLDPREWRDQDPVEVRIARIAIEALEAQAGRTVAADVADAEAGMAALLAAVE 1236
QY 16 P-----DVFAQHQ----- 23
    |||
Db 1237 PPLRGVYHAAGLLDDGLLAHQDAGRLARVLRPKVEGANVHLTLTREQPLDLFVLFSSAG 1296
QY 24 ----- 23
Db 1297 VFGSICQGSYAAGNAFLDALADLRRTOGLAALSIAWGLWAEAGMGSAQRREHESGIWA 1356
QY 24 ----- 23
Db 1357 MPTSRALAAEWLLGTRATORVVIQMDWHAGAAPDASRGRFWDRLVLTATKEASSAVP 1416
QY 24 ----- 23
Db 1417 AVERNRNASVETRYSALVELVRGVVGVMGFTDQGTLDVRRGFAEQGLSLMAVEIRKRL 1476
QY 24 ----- 23
Db 1477 QGELGMLPSLATLFDHPTVERLVEYLLSQALEQDRTDVRSLPATEDPIAIVGAACRF 1536
QY 24 -----EDTERY----- 29
    |||
Db 1537 PGGVEDLESYWQLLTGEGVWVSTVPADRWNGADGRVPGSGEAGQRTYVPRGGFLREVTF 1596
QY 30 ----- 29
Db 1597 DAAPFHSPREMSLDPOQRLLLEVSWEATERAGODPSALRESPTGVFGAGPNEYAERV 1656
QY 30 ----- 29
Db 1657 QELADEAAGLYSGTGNMLSYAAGRLSFFLGLHPTLAVDTACSSSLVALHLCQSLRRGE 1716
QY 30 ----- 29

```

```

Db 1717 CDQALVGVNMLLSPKTFALLSRMHALSPGCRCKTFPSADADGYARAEGCAVVVLKRLSDA 1776
QY 30 ----- 29
Db 1777 QRDRDPILAVIRGTAINHDGSPSSGLTVPSPGAOBALLRQALAHAGVVPADVDFVECHGTG 1836
QY 30 ----- 29
Db 1837 TALGDPIEVRLSDVYGOARPADRPRLILGAAKANLGHMEPAAGLAGLLKAVLALGOEIP 1896
QY 30 ----- 29
Db 1897 AQPELGNLPLLPWEALPVARAAVPMPTDRPRFAGVSSFGMSGTNAHVLEEAPAVE 1956
QY 30 ----- 29
Db 1957 LWPAPERSAELLVLSCKSEGLADAQAARLUREHLDHPELGLGDVAFSLATTRSAMNHLR 2016
QY 30 ----- 29
Db 2017 AVAVTSREGLLAALSAVAQOTPPGAARCIASSSRCKLAFLTGQCAQTGCMGRGLCAAW 2076
QY 30 ----- 29
Db 2077 PAFREAPDRCVALLFDRELDRLPREVMABPGSAESILLDDTQFTOPTALFTVEYALTALWR 2136
QY 30 ----- 29
Db 2137 SWGVEPELVAGHSAGELVAACVAGVFSLEDGVRLVNAARGRLMOGLSAGGMVSLGAPEAE 2196
QY 30 ----- 29
Db 2197 VAAVAHAPHAASVIAAVNGPEQVVIAGVEQAVQAIAGFAARGARTKRLHVSHASHSPLM 2256
QY 30 -----VLTNLN----- 35
    |||
Db 2257 EPMLEFGRVAASVTRYRPSVLSVNSLGVKVADELSPGYVVRHVREAVRFAQGVKALH 2316
QY 36 ----- 35
Db 2317 EAGAGTFVEVGPKPTLLGLLPACLPEAEPTLLASLRAGREAAAGVLEALGRLWAAGSVS 2376
QY 36 ----- 35
Db 2377 WPGVFTAGRRVPLPTYQWQRYWPDIEPDSRRHAAADPTQGWFYRYVDMPEIPRSLQKS 2436
QY 36 ----- 35
Db 2437 EASRGSMLVLADKGVGEAAVAAALSTRGLPCVVLHAPAEATSATAELVTEAGGRSDWQV 2496
QY 36 -----TGAE----- 39
    |||
Db 2497 VLYLWGLDANVVGAEASIDEIGDATRRATAPVLGLARELSTVSCSPRLVVVTRGACIVGDE 2556
QY 40 ----- 39
Db 2557 PAIAPCOAALWGMGRVAALSHPGANGGLVDLDPRASPPQASPIDGEMLVTELLSQETEDQ 2616
QY 40 ----- 39
Db 2617 LAFRHRRHARLVAAPPOGQAAPVLSABASVYLTGGLGGLLIVAQMLVELGARHLVL 2676
QY 40 ----- 39
Db 2677 TSRRGLPDRQAWCEQQPPEIRARIAAVEALEARGARVTVAAVDVADVEPMALVSSVEPP 2736
QY 40 ----- 39
Db 2737 LRGVVHAAGVSVMRPLAETDETLLSVLRPKVAGSMLLHRLHGRPLDLFVLFSSGNAWV 2796
QY 40 ----- 39

```

```
Db 2797 GSHSOGAYAAANAFDGLAHLRRSQSLPALSVAWGLWAEGMADAHAHRLSDIGVLPMS 2856
Qy 40 -----HLV----- 53
Db 2857 TSAALSALORLVETGAARTVTRMDWARFAPVYTARGRRNLLSALVAGRDIIAPSPAAA 2916
Qy 40 ----- 55
Db 2917 TRNWRGLSVAEARVALHEIVHGAVARVLGFLDPSALDPGMGFNEQGLDSLMAVEIRNLQ 2976
Qy 40 ----- 55
Db 2977 AELDVRLSTTLAFDHTVQRLVEHLLVDVLKLEDRSDTQHVRSLASDEPIAIVGAACRFP 3036
Qy 40 ----- 55
Db 3037 GGVEDLESYWOLLAEGVVVAEVPADRMDAADWYDPDPEIPGRYTVTKGAFRLDRLDA 3096
Qy 40 ----- 55
Db 3097 TFRISPREAMSLDPOORLLLEVSWEALESAGIAPOTLRDSPGTVFVGAGPNYYTORLR 3156
Qy 40 ----- 55
Db 3157 GFTDGAAGLYCGTGNMLSVTAGRLSFFLGLHPTLAMDTACSSSLVALHLACQSLRLGEC 3216
Qy 40 ----- 55
Db 3217 DQALVGVGNVLLAPETFEVLLSRMRALSPDGRCKTFSADADGYARGECVAVVLKRLRDAQ 3276
Qy 40 ----- 65
Db 3277 RAGDSLILALIRGSVNHGDPSSGLTVPNGPAQALLRQALSQAGVSPVDVFVECHGTGT 3336
Qy 43 ---DP----- 65
Db 3337 ALGDPIEVOALSEVYGRSGDRPLVLGAAKANVAHLEAASGLASLLKVALLRHQIPA 3396
Qy 45 ----- 65
Db 3397 QPELGELNPHLPNNTLPVAVPRKAVPWGRGARPRRAGVSAFGLSGTNVHVLEEAPEVEP 3456
Qy 45 ----- 65
Db 3457 APAAPARVELVLSAKSAALDAAAARLSAHLSAHPELSLGDVAFSLATTRSPMEHRLA 3516
Qy 45 ----- 65
Db 3517 IATTSREALRGALDAAAQOKTPOGAVRGKAVSRGKLAFLEFTQGGAQMPGMGRGLYETWP 3576
Qy 45 ----- 65
Db 3577 AFREAFDRCVALFDREIDQPLREVWMAAPGLAQOARLDOTAYAQPALFALEYALAAALWRS 3636
Qy 45 ----- 65
Db 3637 WGVEPHVLLGHSIGELVAACVAGVFSLEDAVRLVAARGRLMOALPAGGAWVAIAASEAEV 3696
Qy 45 ----- 70
Db 3697 AASVAPHAATVSTAANGPDVAVVIAGAEVOVLALGATFAARGIRTKRLAVSHAFHSPLMD 3756
Qy 51 ----- 70
Db 3757 PMLEDFORVAATIAYRAPDRPVVSNVTGHVAGPEIATPEYVVRHVRSVAVRFGDGAKALHA 3816
Qy 51 ----- 77
Db 3817 AGAATEVEYGPKEPVLGLLPACLIGEADAVLVPSLRADRSECEVLLAALGANVAVWGALDW 3876
Qy 51 ----- 77
Db 3877 KGVFPDGAARRVALPMYPWQRRHRHMDLTPRSAAPAGIAGRWPFLAGVGLCMCPGAVLHHVLS 3936
```

```
Qy 53 -----HLV----- 55
Db 3937 IGRHPQFLGDHLVFGKVVVPGAFHVAVILSIAAERWPERATIELGTVEFLKAIAMEPDOE 3996
Qy 56 ----- 55
Db 3997 VELHAVLTPEAAGDGYLPFELATLAAPETERRTWTHARORVPTDGPAGALPRLEVLEORA 4056
Qy 56 ----- 55
Db 4057 IQPLDFAGFLDRLSAVRIGWGLMRWLQDGRVGDDEASLATLVPTYPNADHAPLHPILLD 4116
Qy 56 ----- 55
Db 4117 NGFAVSLILSTRSEPEDDGTPLPFAVERVRWRAPVGRVRCGVPRSOAFGVSSFVLVDE 4176
Qy 56 ----- 55
Db 4177 TGEVVAEVEGFCRRRAPREVFLRQESGASTAALYRLDWPPEAPLPDPAERIEESWVVVAA 4236
Qy 56 ----- 55
Db 4237 PGSEMAAALATRLNRCVLAEPKLEAALAGVSPAGVICLWEAGAHEEAPAAAQRVATEGL 4296
Qy 56 ----- 55
Db 4297 SVVQALDRADVRLMWVTMGAVAEGERVQVATAPVMGLGRVTMOERPELSCITLVLEPE 4356
Qy 56 ----- 65
Db 4357 ADAARSADVLLRELGRADDETQVAFRSQKRRVARLVKATTPGEGLLVPPDAESYRLEAGKG 4416
Qy 66 ----- 65
Db 4417 TLDQLRLAPORRAPGGEVEIKVTASGLNPTVLAVLGMYPGDAGPMGDCAGVATAVG 4476
Qy 66 ----- 65
Db 4477 QGVRHVAVDAMVLTGLTHRFVTVDARLVVRQAGLTPAQAAVTPVPAFLTAMLALHDLGN 4536
Qy 66 ----- 65
Db 4537 LRRGERVLIHAAAGGVGMAVQIARWICAEVFATASPSKWAQVQAMGVPRTHIASRTLE 4596
Qy 66 ----- 65
Db 4597 FAETFRQVTGGRVDVVLNALAGEFVDASLSLLSTGGRFLEMGKTDIRDRAAVAAAHPGV 4656
Qy 66 ----- 65
Db 4657 RYRVFDILELAPDRTREILERVVEGFAAGHLRALPVHAFITKAEAAFRFMAQARHOKV 4716
Qy 66 ----- 65
Db 4717 VLLPAPSAAPLAPGTGTVLLTGGLGALGLHVARMLAQQGVPHMVLTRGRGLDTPCAAKAVA 4776
Qy 66 ----- 65
Db 4777 EIEALGARVTIAASDVADRNALEAVLOAIPAEMWLOGVVIHAAGALDDCVLDEOTTDRFSR 4836
Qy 66 ---APNIT----- 70
Db 4837 VLAPKVTGANNLHELHETAGNDLAFVFLPSSMSGLLSAGSQSNYAAANTFLDLAAHRAEG 4896
Qy 71 -----ANLTSSL----- 77
Db 4897 LAAQSLAWGPNWSDGMAAGLSAALQARLARHGMGALSQAQCTALLGQALARPETOLGAMS 4956
Qy 78 ----- 77
Db 4957 LDVRAASQASGAAVPPVWRALVRABARHTAAGAQAALARLALGALPEARRADEVRKVVQAE 5016
```



```
QY 78 ----- 77
Db 5017 IARVLSWASAAPVDRPLSDGLDLSLTAVELRNVLGQVATLPATLAFDHPTVDALTR 5076
QY 78 ----- 77
Db 5077 WLLDKVLAABSVSSAKSSPOVALDEPIAIIIGICRPPGGVADPESFWRLLBEGSDAVV 5136
QY 78 ----- 77
Db 5137 EVPHERWDIDAFYDPDPDVRGKMTTRFCGFLSDIRDPDPAFFGISPREATTMDPQORLLL 5196
QY 78 ----- 77
Db 5197 ETSWEAFERAGILPERLMGSDTVGVGLFYOEYAAAGGIEAFDGYLGTGTASVAGRI 5256
QY 78 ----- 77
Db 5257 SYVILGKGPSLTVDTACSSSLVAVHLACQALRRGECSSVALAGGVALLMTPTATVEFSRLR 5316
QY 78 ----- 77
Db 5317 GLAPDGRCKSFSAADGVHSEGCMALLKPLRDAQRDGDPILAVIRGTAVNQDGRSNGL 5376
QY 78 ----- 77
Db 5377 TAPNGSSQOEVIIRALEQAGLADPADSVYVECHGTGTTGLDPIEQALGAVLAQGRPSDRP 5436
QY 78 ----- 77
Db 5437 LVIGSVKSNIGHTQAAAGVAGVIKVALALBERGLIPRSLHFDAPNHPIMSELAVQVAAKP 5496
QY 78 ----- 77
Db 5497 VENTRNGVPRRAGVSFGVSGTNAHVHLEAPAAAFAPAAARSaelFVLSAKSAAALDAQ 5556
QY 78 ----- 77
Db 5557 AARLSAHVAHPGLGDLAFSLATTRSPMTYRLAVAATSREALSAAALDAAQOQAPPA 5616
QY 78 ----- 84
Db 5617 ARCHASTGNAPKVVVFPPGGGQWLGOKILLSEEPVFRDALSAACDRAIOAEGWLSLAE 5676
QY 85 ----- 84
Db 5677 LAADETTSQLRIDVVQPALFAIEVALSALWRSWGEVDPDAVVGHSWGEVAAAHVAGALS 5736
QY 85 ----- 84
Db 5737 EDVAIIICRRSILLRRISQGGEMAVVELSLAEAEALIGYEDRLSVAVSNSPRSTVLAGE 5796
QY 85 ----- 84
Db 5797 PAALAEVLAILAAGVFCRRVKVDVASHSPQIDPLRDELLAALGELEPRQATVSMRSTVT 5856
QY 85 ----- 84
Db 5857 STIMAGPELVASYADNVRQPVFAEAVQSLMEDGHGLFVEMSPHILTTTSVEIRATK 5916
QY 85 ----- 84
Db 5917 REGVAVGSLRRCODERLSNLEALGALVHVGQAVGWERLFSAGGAGLRRVPLPTYPWQORER 5976
QY 85 ----- 90
Db 5977 YWVDAPTGAAGGRFAHAGSHPLIGEMQTLSTQRSTRVWETTLDLKELPWLDHRVQGA 6036
QY 91 ----- 94
Db 6037 VVFFCAVLEMALSSGAEALGGPLQVSDVVLAELAFADDTPAAVQVMATEERPGRLOF 6096
QY 95 ----- PGH ----- 97

|||
Db 6097 HVASRVPGGGAARSHARGVLRQIERAEVPARLDLAALARLOASAPAAATYAALAEWG 6156
QY 98 ----- 97
Db 6157 LEYGPAFOGLVELWRGEGBALGRVRLPEAAGSPAACRLHPALLDACFHVSSAFADRGBAT 6216
QY 98 ----- 97
Db 6217 PWVPVEIGSLRWFORPSGELWCHARSVSHGKPTPDRRSTDFMVVDSTGAIVAEISGLVAQ 6276
QY 98 ----- 97
Db 6277 RLAGVRRREDDFMPEPAWEPTAVPGSEVMAGRMLLIGSGGLGAALHSALTAEAGHSV 6336
QY 98 -----AD 99
Db 6337 HATGRGTSAGLQALLTASFDGOAPTsvVHLGSLDERGVLDADAPFDADALEESLVRGCD 6396
QY 100 LVLY-----|TR-----|
|||
Db 6397 SVLWTVQAVAGAFRDPRLWLVRGAQAIAGAGDVSVAQAPLLGLGRVIALEHAELRCAR 6456
QY 107 ----- 106
Db 6457 IDLPARRDGEVDELLAELLADDAEEVAFRGGERRVARLVRRLPETDCREKIEPAEGRP 6516
QY 107 FDLEL-----|
|||
Db 6517 FRLEIDGSGVLDLVRATERPPGPCEVEIAVEAAGLNFLDVMRANGIYPGGDGPVAL 6576
QY 112 ----- 111
Db 6577 GABCSGRIVAMGEGVESLRIGODVAVAPFSFGTHVTIDARMLAPRPAALTAQAAALPV 6636
QY 112 ----- 111
Db 6637 AFMTAWGLVHLGRLAGERVLIHSATGGTGLAAVOIARHLGAEIFATAGTPEKRAWLRE 6696
QY 112 -----PGN----- 115
Db 6697 QGIAHVMSRSLDFAEQVLAATKGEVDVVLNSLSGAAIDASLSTLVPDGRFIELCKTDI 6756
QY 116 ----- 115
Db 6757 YADRSGLAHFRKSLSYSADVLAGLAVRRPERVAALLAEWOLLARGALQPLPFIPLS 6816
QY 116 ----- 115
Db 6817 RAADAFKVAQAQHLGKLVLALEDPDVRIrVPGESGVAIRADGAYLVTTGGLGGLSVAG 6876
QY 116 -----ROV--- 118
Db 6877 WLAEOGAGHLVLVGRSGAVSAEQQTAVAALAEHARGVTVARADVADRAQMERILREVTA 6936
QY 119 ---RGVTQGG---|
|||
Db 6937 GMPLRGVVHAAGILDDGLLMQOTPARFRVAVMAPKVRGALHHLHALTREATPLSFVLYASGA 6996
QY 127 -----ACSPTWSCL----- 135
Db 6997 GLLGSPGQGNAAAANTFLDALAHHRRAOGLPALSIDMGLFADVGLAAGQONRGARLVTRG 7056
QY 136 ---ITEDTG---|
|||
Db 7057 TRSLTPDEGLWALERLLDGRDTRQAGVMPFDVRQWVEFYPAASRRSLRLMTARRVASGR 7116
QY 144 ----- 143
Db 7117 LAGDRDLERLATAEAGARAGMLQEVVRAQVSVLRSEGLVDVDAPLTSLGMSLMLGLE 7176
QY 144 -----LGVTI 148
|||:
```

Db 7177 LRNRIEAVLGITM 7189

RESULT 10

US-09-568-486-5
 ; Sequence 5, Application US/09568486
 ; Patent No. 6355459
 ; GENERAL INFORMATION:
 ; APPLICANT: Schupp, Thomas
 ; APPLICANT: Ligon, James
 ; APPLICANT: Molnar, Istvan
 ; APPLICANT: Zirkle, Ross
 ; APPLICANT: Cyr, Devon
 ; APPLICANT: Goerlach, Joern
 ; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
 ; FILE REFERENCE: 4-30582A
 ; CURRENT APPLICATION NUMBER: US/09/568,486
 ; CURRENT FILING DATE: 2000-05-10
 ; PRIOR APPLICATION NUMBER: 09/335,409
 ; PRIOR FILING DATE: 1999-06-17
 ; NUMBER OF SEQ ID NOS: 30
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 5
 ; LENGTH: 7257
 ; TYPE: PRT
 ; ORGANISM: Sorangium cellulosum
 US-09-568-486-5

Query Match 27.0%; Score 209; DB 4; Length 7257;
 Best Local Similarity 1.5%; Pred. No. 3.1;
 Matches 93; Conservative 23; Mismatches 32; Indels 6165; Gaps 30;
 1 AAGG-----4
 ||||
 877 AAGGSWSGVFTAGRRVPLTPYQRYWIEAPAEGLGATAADALAQWFRVDWDEM 936
 5 -----ILH-----7
 937 PRSSVDSRRASGWLVLADRGVGEAAALSSQGCSCAVLHAPAEASNAVEQVTOALG 996
 8 -----7
 997 GRNDWQGVLYLWGLDAVVEAGASAEVAKVTHLAAAPVLALIOALGTGPRSPRLWITVRG 1056
 8 -----7
 1057 ACTVGGEPDAAPCAALMGWRVALEHPGSGWGLVDLPDPESPTEVEALVAELLSPDAE 1116
 8 -----7
 1117 DQLAFQRRRAARLVAAPEEGNAAPVLSAEGSYLTGGLGALGLVARMLVERGAGHL 1176
 8 -----LE-----LLVAVG 15
 1177 VLISRHGLPDREWGGRDQPEVVRARIAATEALEAOGARTVVAADVADAEAGMAALLAAVE 1236
 16 P-----DVFOAHQ-----23
 1237 PPLRGVVAAGLLDDGLLAHQDAGRLARVLRPKVEGAWVLHTLTREQPLDLFLVLFSSAG 1296
 24 -----23
 1297 VFGSIGQSYAAGNAFLDALADLRRTOGLAALSIAWGLWAEAGMGSGAQRREHEASGIWA 1356
 24 -----23
 1357 MPTSRALANWELLGTRATORVVIQMDWAHAGAAPRDAASGRFRWDLVTATKEASSAVP 1416
 24 -----23
 1417 AVERWRNASVETRSALVELVRGVAGVNGFTDQGTLDVYRGRFAEQGLDLSLMAVEIRKEL 1476
 24 -----23

Db 1477 QGELCMPLSATLAFDHPPTVERLVEYLSQALELQDRTDVRSLPATEDPDAIACAACRF 1536
 QY 24 ----EDTERY-----29
 ||||
 Db 1537 PGGVEDLESYQQLLTEGVVSTEVPADRWNGADGRVPGSGEAOQOTYVPRGGFLREVETFF 1596
 QY 30 -----29
 Db 1597 DAAFFHISPREAMSLDPOORLLLEVSWEAIERAGQDPSALRESPTGVFVGAGPNEYAERV 1656
 QY 30 -----29
 Db 1657 QELADEAAGLYSGTGNMLSVAAGRLSPFLGLHGPTLAVDTACSSSLVALHLCQSLRGE 1716
 QY 30 -----29
 Db 1717 CQALVGVGNMLLSPKTFALLSRMHALSPGCRCKTFSADADGYARAEGCAVVVLKRLSDA 1776
 QY 30 -----29
 Db 1777 QRRDPILAVIRGTAINHDGSPSGLTVPSGPAQBALLRQALAHAGVVPADVDFVECHGTG 1836
 QY 30 -----29
 Db 1837 TALGDPIEVRALSDVYGOARPADRPRLILGAANKLGHMEPAAGLAGLLKAVLALGOEQIP 1896
 QY 30 -----29
 Db 1897 AQPELGEINPLLPWEALPVAVARAAPWPRTRDRPRFAGSVSGMGTNAHVLEADPAVE 1956
 QY 30 -----29
 Db 1957 LWPAAPERSAELLVLSCSEGLDAQARLREHLDMHPELGLDGVAFSLATTRSAMNHRL 2016
 QY 30 -----29
 Db 2017 AVAVTSREGLLAALSAVAOQTTPGAARCIASSSRGKLAFLFTQOQAOTPCMGRGLCAAW 2076
 QY 30 -----29
 Db 2077 PAFREAFDRCVALFDRELDRLREVMMWAEPSAESLLDDQTAFTOPALFTVEYALTALWR 2136
 QY 30 -----29
 Db 2137 SWGVEPELVACHSAGELVAACVACVPSLEEDGVRLVAARGRLMOGLSAGGAMVSLGAPEAE 2196
 QY 30 -----29
 Db 2197 VAAAVAPHAASVSIAAVNGPEQVVIAGVEQAVQAIAGFAARGARTKRLVSHASHSPLM 2256
 QY 30 -----VLTNLN-----35
 2257 EPMLEEFGRVAASVTYRRRPSVLSVNSLGSKVVADELSPGYWVRHVREAVRFADGVKALH 2316
 QY 36 -----35
 Db 2317 EAGAGTFVEVGPKPTLLGLLPACLPPEAEPTLLASLRAGREAAAGVLEALGRLWAAGSVS 2376
 QY 36 -----35
 Db 2377 WPGVFTAGRRVPLTPYQRYWPDIEPDSRRHAAADPTQGMFYRVDMWPEIPRSLQKS 2436
 QY 36 -----35
 Db 2437 EASRGSWLVADKGGVGEAAVAAALSTRGLPCVVLHAPAEATSATAELVTEAAGRSQWQV 2496
 QY 36 -----IGAE-----39
 2497 VLYLWGLDAVGAESAIDEIGDATRRATAPVLGLARFLSTVSCSPRLVVVTRGACIVGDE 2556
 QY 40 -----39

```
Db 2557 PAIAPQAAALMGMRVAALHEHFGAWGLVDLDPRASPPQASPIDGEMLVTELLSQETEDQ 2616
Qy 40 -----
Db 2617 LAFRHRHRAARLVAAAPGQAAAPVLSAEASYLVGTGGLGLGLIIVAQMLVELGARHLVL 2676
Qy 40 -----
Db 2677 TSRRGLPDRQAWCEQPPEIRARIAAVEALEARGARVTVAAVDVADVEPM TALVSSVEPP 2736
Qy 40 -----
Db 2737 LRGVVHAAGVSVMRPLAETDETLLSVLRPKVAGSWLLHLLHGRPLDLFLVLFSSGAAMV 2796
Qy 40 -----
Db 2797 GSHOGAYAAANAFDGLAHLRRSQSLPALSVAWGLWAEGMADAEAHARLSDIGVLPMS 2856
Qy 40 -----
Db 2857 TSAALSALORLVETGAAQRTVTRMDWARFAPVYTARGRRNLLSALVAGRDIIAPSPAAA 2916
Qy 40 -----
Db 2917 TRNWRGLSVAEAREVALHEIVHGAVARVLGFLDPSALDPMGFNEOGLDSLMAVEIRNLLQ 2976
Qy 40 -----
Db 2977 AELDVRLSTTLAPDHTVORLVEHLLVDVLKLEDRSDTQHVRSLASDEPIALVGACRFP 3036
Qy 40 -----
Db 3037 GGVEDLESYWQLLAEGVVVAEVPADRMDAADWYDPPEIPGRYTVTKGAFLDLQLRDLA 3096
Qy 40 -----
Db 3097 TFFRISPREAMSLDPOORLLLEVSWEALESAGIAPDTRDSTGTVFVGAGNEYTORLR 3156
Qy 40 -----
Db 3157 GFTDGAAGLYGTGNMLSVTAGRLSFFLGLHGPTLAMDTACSSSLVALHLACQSLRLEC 3216
Qy 40 -----
Db 3217 DOALYGVNVLLAPETFFVLLSNRMLSPDGRCKTFSADADYARGECAVVYLKRLDAQ 3276
Qy 40 -----
Db 3277 RAGDSILALIRGSAVNHDPSSGLTVPNGPAQALLRQALSQAGVSPVDVDFVECHGTGT 3336
Qy 43 ---DP-
|||
Db 3337 ALGDPFIEQALSEVYGPGRSGDRPLVLGAAKANVAHLEAASGLASLLKAVLALRHEQIPA 3396
Qy 45 -----
Db 3397 OPELGELNPLPMNTLPVAVPRKAVPNRGARPRRAGVSAFGLSGTNVHVHVEEAPEVEP 3456
Qy 45 -----
Db 3457 APAAPARPELVVLVLSAKSAALDAAAARLSAHLSAHPSELSDGVAFSLATTRSPMEHRLA 3516
Qy 45 -----
Db 3517 IATTSREALRCLDAAAQKQTPQAVRGKAVSSRGKLAFLFTGGAQPMGMGRGLYETWP 3576
Qy 45 -----
Db 3577 AFREAFORCVALEFDREIDOPREVMMAAPGLAQARLDQTYAQPALPALEYALAAALWRS 3636
Qy 45 -----
Db 3637 WGVPEPHVLLGHSIGELVAACVAGVFSLEDAVRLVAARGRLMOALPAGGAMVAIAASEAEV 3696
Qy 45 -----
Qy 45 -----SLGAOF----- 50
|||
Db 3697 AASVAPHAATVSYIAAVNGPDVAVIAGAEVQVLALGATTAARGITKRLAVSHAPHSPMD 3756
Qy 51 -----
Db 3757 PMLEDFORVAATIAIRAPDRPVVSNVTGHVAGPEIATPEYVWRHVRSAVRFGDGAKALHA 3816
Qy 51 -----
Db 3817 AGAATFVEVGPKPVLLGLLPACLGADAVLVPSLRDRSECEVVLAALGAWYAWGGALDW 3876
Qy 51 ---RV-
|||
Db 3877 KGVPDGGARRVALPMYPMQORERHMDLTPRSAAPAGIAGRWPLAGVGLCMFPGAVLHHVLS 3936
Qy 53 ---HLV-
|||
Db 3937 IGPRHQPFGLGDHLVFGKVVPVPGAFHVAVILSIAAERMPRAIBLTGTVEFLKATAMEPDOE 3996
Qy 56 -----
Db 3997 VELHAVITPEAAGDGYLFLATLTAAPETERRWTHARGRVQPTDGPAGPALPRLEVLEDR 4056
Qy 56 -----
Db 4057 IQPLDFAGFLDLRSVIRIGWPLRWLQDGRVGDSEASLATLVPTYNAHDVAPLHPILLD 4116
Qy 56 -----
Db 4117 NGFANVLLSTRSEPEDDGTPLPFAVERVVRWRAPVGRVRCGVPRSQAFVSFVLVDE 4176
Qy 56 -----
Db 4177 TGEWVAEVEGFCRRAPREVFLRQESGASTAALYRLDWPEAPLPDAPAEIEESWVVA 4236
Qy 56 -----
Db 4237 PGSEMAALATRLNRCVLAEPKGLEAALAGVSPAGVICLWBAGAEHEEAPAAAQRVATEGL 4296
Qy 56 -----
Db 4297 SVVQALRDRAVRLMWTVGAVAVAGERVQVATAPVWGLGRVTHQERPELSCTLVLEPE 4356
Qy 56 ---KXVILTEPEG-
|||
Db 4357 ADAARSADVLLRELGRADDETQVAFRSGKRVRARLVKATTPGGLLVPAESYRLEAQKG 4416
Qy 66 -----
Db 4417 TLDQLRLAPAORRAPPGGEVEIKVTASGLNFRTVLAVLGMYPGDGPMGGDCAGVATAVG 4476
Qy 66 -----
Db 4477 QGVRHVAVDAMTGLTLHRFTVDARLVVRQAPAGLTPAQAAVTPVAFVTAWLALHDLGN 4536
Qy 66 -----
Db 4537 LRRGERVLIHAAAGGVGAAAVQIARMTIGAEVFATASPSKWAAVQAMGVPRTHIASRRTLE 4596
Qy 66 -----
Db 4597 FAETFRQVTGGRGVVDVNLNAGEFVDASLSLLSTGGRFLEMKGKDIRDRAAVAAAHPGV 4656
Qy 66 -----
Db 4657 RYRVFDILELAPORTREILERVVEGFAAGHLRALPVHAFATKAEAAFRMAQARHQKV 4716
Qy 66 -----
Db 4717 VLLPAPSAAPLAPTGTVLLTGGLGALGLHVARWLAAQGVPHMVLTGRRGLDTPGAARAVA 4776
```

```
QY 66 ----- 65
Db 4777 EIEALGARVTIAASDVADRNALAEAVLQAI PAEWPLQGVIIHAAGALDDGVLDQTTDFSR 4836
QY 66 --APNIT----- 70
Db 4837 VLAPKVTGAWNLHETAGNDLAFVFLPSSMSGLLGSAGQSNYAAAANFTLDALAAHRAEG 4896
QY 71 -----ANLTSSL----- 77
Db 4897 LAAQSLAWGPMWSDGMAAGLSAALQARLARHMGALSPAQCTALLGQALARPETQLGAMS 4956
QY 78 ----- 77
Db 4957 LDVRAASQASGAAPVPPVWRALVPAEAREHTAAGAGALAAALGALPEARRADEVKVVQAE 5016
QY 78 ----- 77
Db 5017 IARVLSWSAASAVPDRPLSDGLDSLTAVELRNLVQRVGATLPATLAFDHTVDALTR 5076
QY 78 ----- 77
Db 5077 WLLDKVLAVAEPSVSSAKSPQVALDEPIAIIIGICRFPQGVADPESFMRLLLEEGSDAVV 5136
QY 78 ----- 77
Db 5137 EVPHERNIDAFYDPPDVRGKMTTRFGGFLSDIDRDPAPFGISPREATTMPQORLLL 5196
QY 78 ----- 77
Db 5197 ETSWEAFERAGILPERLMGSDTVGVGLFYQEVAAALAGGIEAFDGYLGTTGTASVASGRI 5256
QY 78 ----- 77
Db 5257 SYVLGLKGPSLTVDTACSSSLVAHLACQALRRGECVSALAGGVALMLTPATFVEFSRLR 5316
QY 78 ----- 77
Db 5317 GLAPDGRCKSFAAADGVGSEGCMALLKPLRDAQRDGDPILAVIRGTAVNQDGRSNGL 5376
QY 78 ----- 77
Db 5377 TAPNGSSQOEVIIRALEOAGLADVPVSVVECHGTGTTGLDPIEVQALGAVLAQGRPSDRP 5436
QY 78 ----- 77
Db 5437 LVIGSVKSNIGHTQAAAGVAGVIKVALLALERGLIPRSLHFDAPNPHI PWSELAVQVAAKP 5496
QY 78 ----- 77
Db 5497 VEMTRNGVPRRAGVSGVGTNAHVVLEBAPAAAFAPAAARSANELFVLSAKSAAALDAQ 5556
QY 78 ----- 77
Db 5557 AARLSAHVVAHPELGLGDLAFSLATTRS PMTYRLAVAATSREALSAAALDTAAQGOAPPA 5616
QY 78 -----GWS----- 84
Db 5617 ARGHASTGSAPKVVFPFGOGSOWLCMGOKLLSEEPVFRDALSACDRAIQAEAGHSLAE 5676
QY 85 ----- 84
Db 5677 LAADETTSQLGRIDVVQPALFAIEVALSALWRSGVPEPDVVVGHSGMGEVAAAHAHAGALS 5736
QY 85 ----- 84
Db 5737 EDVAIIICRRSLLRRISGEMANVVELSLAEAEALGYEDRSLVAVSNSPRSTVLAGE 5796
QY 85 ----- 84
Db 5797 PAALAEVLAIIAAKGVFCRRVKVDVASHSPQIDPLRDELLAALGELEPRQATVMSRSTVT 5856
QY 85 ----- 84
```

```
Db 5857 STIMAGPELVASYWADNRQPVRFAEAVQSLMEDCHGLFVEMSPHPILTTSVTEEIRATK 5916
QY 85 ----- 84
Db 5917 REGVAVGSLRRQODERLSMLEALGALWVHGOAVGWERLFSAGGAGLRVRPLPTYPWORER 5976
QY 85 -----QTINPE----- 90
Db 5977 YWVDAPTGAAGGSRFAGSHPLLGEMQTLSTQRSTRVWETTLDLKRLPMLGDHRVGA 6036
QY 91 -----DDTD----- 94
Db 6037 VVFPGAAYLEMALSGAEALGGPLOVSDVVLAELAFADDTTAAVQVMATEERPGRLOF 6096
QY 95 -----PGH----- 97
Db 6097 HVASRVPGHGGNAFNRSHARGVLRQIERAEVPARLDLALRARLARQASAPAAATYAALAE 6156
QY 98 ----- 97
Db 6157 LEYGPAFQGLVELWRGEALGRVRLPEAAGSPAACRLHPALLDACFHVSSAFADRGAT 6216
QY 98 ----- 97
Db 6217 PWVPVEIGSLRWFORPSGELWCHARSVSHGKPTPDRRSTDFWVVDSTGAIVAEISGLVAQ 6276
QY 98 ----- 97
Db 6277 RLAGVRRREDDWFMPEPAWETA VPGSEVMAGRMLLJGSGGLGNALHSALTAEAGHSV 6336
QY 98 -----AD 99
Db 6337 HATCRGTSAAGLQALLTASFDGQAPTSVVHLGSLDERGVLDADAPFDADALEESLRGCD 6396
QY 100 LVLY-----ITR----- 106
Db 6397 SVLTVQAVAGAFRDPRLWLVTGTAQATGAGDVSAQAAPLLGLGRVIALEHAEARCAR 6456
QY 107 ----- 106
Db 6457 IDLPARRDGEVDDELLAEALLADAEERVAFRGGERRVARLVRLPETDCREKIEPABGRP 6516
QY 107 FDLEL----- 111
Db 6517 FRLEIDGSGVLDLVLRA TERRPPGGEVEIAVEAAGLNFLDVWRAMGIYPGPGDGPVAL 6576
QY 112 ----- 111
Db 6577 GAECSGRIVANGEVESLRIGQDVVAVAPFSFGTHVTIDARMLAPRPAALTAQAQAALPV 6636
QY 112 ----- 111
Db 6637 AFMTAWYGLVHLGRLRAGERVLIHSATGGTGLAAVOIARHLGAEIFATAGTPEKRAWLRE 6696
QY 112 -----PGDN----- 115
Db 6697 QGIAHVMSRSLDFAEQVLAATKGEVDVVVLSLSGNAIDASLSTLVPDGRFIELGKTDI 6756
QY 116 ----- 115
Db 6757 YADRSGLAHFRKSLSYSAVDLAGLAVRRPVERVAALLAEVVDLLARGALQPLPVEIFPLS 6816
QY 116 ----- 115
Db 6817 RAADAFKMAQAOHLKGLVLALEDPPVRI RVPGESGVATRADGAYLVLTGGLGGLGSVAG 6876
QY 116 -----ROY----- 118
Db 6877 WLAEOQAGHLVLYGRSGAVSAEQOTAVAALEAHGARVTVARADVADRAQMERILREVITAS 6936
QY 119 -----RGVTQLGG----- 126
```

Db 6937 GMPLRGVVHAAGILDDGLLMOOTPARFRAVWAPKVRGALHLHALTREAPLSFFVLYASGA 6996
Qy 127 -----ACSTWCL----- 135
Db 6997 GLLGSPGQGYAAANTFLDALAHRRRAOGLFALSIDWGLFADVGLAAGQQNRGARLVTRG 7056
Qy 136 ---ITEDTG---FD----- 143
Db 7057 TRSLTPDEGLWALERLLDGRDRTQAGWMPFDVRQWVEFYPAASRRRLSRLMTAPRVASGR 7116
Qy 144 ----- 143
Db 7117 LAGDRDLLERLATAEAGARAGMLQEVVRAQVSVLRSEGKLDVDAPLTSLGMDSLMGL 7176
Qy 144 -----LGVTI 148
Db 7177 LNRRIEAVLGITM 7189
RESULT 11
US-09-568-472-5
; Sequence 5, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-5
Query Match 27.0%; Score 209; DB 4; Length 7257;
Best Local Similarity 1.5%; Pred. No. 3.1;
Matches 93; Conservative 23; Mismatches 32; Indels 6165; Gaps 30;
Qy 1 AAGG----- 4
Db 877 AAGGSVMPGVFPPTAGRRVPLTPYQQRVWIEAPAEGLGATADALAQNWFYRVDWPEM 936
Qy 5 -----ILH----- 7
Db 937 PRSSVDSRRARRSGGWLVLADRGVGEAAAAALSSQGCSCAVLHAPAEASVAEQVTOALG 996
Qy 8 ----- 7
Db 997 GRNDHQGVLYLWGLDVAVEAGASAEVAKVTHLAAAPVLALQALGTGPRSLWIVTRG 1056
Qy 8 ----- 7
Db 1057 ACTVGEPPDAAPCOALWGMGRVAALHPGWSGGLVDLPDEEPSPEVEALVAELSPDAE 1116
Qy 8 ----- 7
Db 1117 DOLAFGRRRRAARLVAAPPEGNAAPVLSAEGSYLVGTGGLGALLVARLWVERGAGHL 1176
Qy 8 -----LE-----LLVAVG 15
Db 1177 VLISRHGLPDREWGRDQPEVRARIAIEALEAQGARVTVAADVADAEAGMAALLAAVE 1236
Qy 16 P-----DVFOAHO----- 23

Db 1237 PPLRGVYVHAAGLLDDGLLAHADAGRLARVLRPKVEGAWVLHTLTREOPLDLFVLPSSASG 1296
Qy 24 ----- 23
Db 1297 VFGSICQSSYAAGNAFLDALADLRRTOGLAALSIAWGLWAEGBGMSQORREASGIWA 1356
Qy 24 ----- 23
Db 1357 MPTSRALAAMEWLLGTRATORVWIQMDWAHAGAAPRDSRGRFMDRLVTATKEASSAVP 1416
Qy 24 ----- 23
Db 1417 AVERWRNASVETRSALYELVGVVAGVMGFTDQGLDVRGFAEOGLSLMAVEIRKRL 1476
Qy 24 ----- 23
Db 1477 QGELGMLSATLAFDHPVTERLVYLLSQALELQDRDTRVRSVRLPATEDDPIAIVGAACRF 1536
Qy 24 ---EDTERY----- 29
Db 1537 PGGVEDLESYWOLLTEGVVSVTEVPADRMNGADGRVPGSGEAGQRTQVVPGGFLREVET 1596
Qy 30 ----- 29
Db 1597 DAAPFHISPREAMSLDPQORLLLEVSWEAIERAGQDPSALRESPTGVFVAGSNEYAERV 1656
Qy 30 ----- 29
Db 1657 QELADEAAGLYSGTGNMLSVAAGRLSFFLGLHCPITLAVDTACSSSLVALHLGCOSLURR 1716
Qy 30 ----- 29
Db 1717 CDQALYGGVNMLLSPKTFALLSRMHALSPGGRCKTFADADGYARAEGCAVVVLRKLSA 1776
Qy 30 ----- 29
Db 1777 QRDRDPILAVIRGTAINHDGSSGLTVPSGPAQEALLRQALAHAGVVPADVDFVECHGTG 1836
Qy 30 ----- 29
Db 1837 TALGDPFIEVRALSDVYGOARPADRPLILGAANKILGHMEPAAGLAGLLKAVLALGOEIP 1896
Qy 30 ----- 29
Db 1897 AQPELGELNPLLPWEALPVAVARAAPVWPRTDRPRFAGVSSFGMSGTNAHVLEAPAVE 1956
Qy 30 ----- 29
Db 1957 LWPAAPERSAELLVLSKSEGALDAQARLREHLDHMPHGLGDLVAFSLAETRSMNHRL 2016
Qy 30 ----- 29
Db 2017 AVAVTSREGLLAALSVAAGQTPFGAARCIASSSRGKLAFLTGQGAOTPGMGRGLCAAW 2076
Qy 30 ----- 29
Db 2077 PAFREAFDRCVALFDRELDRLPREVMWAEPCGSABESLLLDOTAFQPALFTVEYALTALWR 2136
Qy 30 ----- 29
Db 2137 SWGVEPELVAGHSAGELVAACVAGVPSLEDDVRLVAARGRLMOGLSAGGANVSLGAPEAE 2196
Qy 30 ----- 29
Db 2197 VAAAVAPHAASVSTAANGPEQVVIAGVEQVAVOIAAGFAARGARTKRLHVSHASHSPLM 2256
Qy 30 -----VLTNLN----- 35
Db 2257 EPMLEEFGRVAASVTYRRPSVSLVNSLCKVVADELSAPGYWVRHVREAVRFADGVKALH 2316
Qy 36 ----- 35

Db	2317	EAGAGTFVEVGKPTLLGILLPACLPENAEPTLLASLRAGREAAAGVLEALGRLWAAGGSVS	2376
QY	36	-----	35
Db	2377	WFCVFPTAGRRVPLPTYPHQRYWPDIEPDSRRHAADPTQGWFYRDWPEIPRSLOKS	2436
QY	36	-----	35
Db	2437	EEASRGSWLVLADKSGVGEVAAALSTRGLPCVVLHAPETSATAELVTEAAGGRSDQV	2496
QY	36	IGAE :	39
Db	2497	VLYLGLDVAWGAEASIDEIGDATRRATAPVLGLARFLSTVSCPSRLWVTRGACIVGDE	2556
QY	40	-----	39
Db	2557	PAIAPCOAALMGMRVAALEHPGAWGLVDLDPRASPPQASPIDGEMLVTELLSOETEDQ	2616
QY	40	-----	39
Db	2617	LAFRHRRHARLVAAPPGQONAPVLSAEASYLVGTGGLGLLIVAQWLVELGARHLVL	2676
QY	40	-----	39
Db	2677	TSRRGLPDRQWCEQPPEIRARIAAEALEARGARTVVAADVADVEPMTALVSSVEPP	2736
QY	40	-----	39
Db	2737	LRGVVHAAGVSMRPLAETDETLLESVLPRKVAGSWLLHRLHGRPLDLFLVSSGAAMV	2796
QY	40	-----	39
Db	2797	GSHSGQAAANAFLDGLAHLRRSQSLPALSVAWGLWAEGMADAAHARLSDIGVLPMS	2856
QY	40	-----	39
Db	2857	TSAALSALORLVETGAAQRTVTRMDWARFAPVYTARGRNLLSALVAGRDIIAPSPPAAA	2916
QY	40	-----	39
Db	2917	TRNWRGLSVAERVALHEIVHGAVARVLGFLDPSALDPGMGFNEQGLDSLMAVEIRNLLO	2976
QY	40	-----	39
Db	2977	AELDVRLSTTLAFDHTVQRLVEHLVDVLKLEDRSDTQHVRSLASDEPIAIVGAACRFP	3036
QY	40	-----	39
Db	3037	GGVEDLESYQWLLAEGVVVSAEVPADRWDADWDYDPEIPGRTYVTKGAFRLDLQLDA	3096
QY	40	-----	39
Db	3097	TFFRISPREAMSLDPOQRLLEVSWEALESAGIAPDTRLDSPTGVFVGAGPNEYTQRLR	3156
QY	40	-----	39
Db	3157	GFTDGAAGLYGGTGNNMLSVTAGRLSPFLGLHGPTLAMDTACSSSLVALHLCQSLRLGEC	3216
QY	40	-----	39
Db	3217	DOALVGVNVLLAPETVLLSRMRLSPDGRCKTFSADADGARGEGCAVVVLKRLRDAQ	3276
QY	40	LUR :	42
Db	3277	RAGDSLALIRGSVNHDPSSGLTVPNGPAQOALLRQALSQAGVSPVDVDFVECHGTGT	3336
QY	43	DP :	44
Db	3337	ALGDPIEQVQALSEVYGFGRSGDRPLVLGAAKANVAHLEAASGLASLLKAVLALRHEQIPA	3396
QY	45	-----	44
Db	3397	QPELGINPHLPWNTLPAVAPRKAVPWGRGARPRRAGVSAFGLSGTNVHVLEAPEVEP	3456
QY	45	-----	44
Db	3457	APAAPARPVELVLSAKSAAALDAAAARLSAHLSAHPELSGLDVAFSLATTRSPMEHRLA	3516
QY	45	-----	44
Db	3517	IATTSREALRGALDAAAOQKTPQGAVRGKAVSSRGKLAFLPTGOGAQMPCMGRLYETWP	3576
QY	45	-----	44
Db	3577	AFREAPDRCVALLFDREIDOPUREVMMAPGLAQAAARLDOTAYAPALFALEYAALWRS	3636
QY	45	-----	44
Db	3637	MGVEPHVLLGHSIGELVAACVAGVFSLEDAVRLVAARGRLMQALPAGGANVAIAASEAEV	3696
QY	45	SIGAOF :	50
Db	3697	AASVAPHAATVSIAAVNGPDVAVIAGAEVQVLGATPAARGIRTKRLAVSHAFHPLMD	3756
QY	51	-----	50
Db	3757	PMLEDFORVAATIAYRAPDRPVSVNVTHVAGPEIATPEYVVRHVRSVAVRFGDGAKALHA	3816
QY	51	-----	50
Db	3817	AGAATFVEVGPKVLLGLLPACLGADAVLPSURADRSECEVVLAALGAWYANGGALDW	3876
QY	51	RV :	52
Db	3877	KGVPDGGARRVPLWPMQORERHMDLTPRSPAAGIAGRWPLAGVGLCMPGAVLHHVLS	3936
QY	53	HLV :	55
Db	3937	IGPRHPFLGDLHVLFGVWVPGAVHVAVILSIAERWPERAIBLTGVFEFLKATAMEPDQE	3996
QY	56	-----	55
Db	3997	VELHAVLTPTEAGDGYLFELATLAAPETERRTTHARGVQPTDGPALPRLEVLDR	4056
QY	56	-----	55
Db	4057	IQPLDFAGFLDRLSAVRIGHGLWRWLQDGRVGDSEASLATLVPTYNAHDVAPLHPILDD	4116
QY	56	-----	55
Db	4117	NGFAVSLSTRSEPDGDTPLPFAVERVRWRAPVGRVRCGGVPRSQAFGVSSFVLVDE	4176
QY	56	-----	55
Db	4177	TGEVVAEVEGFVCRAPREVFLROESGASTAALYRLDWPPEALPDAPAERIEESWVVAA	4236
QY	56	-----	55
Db	4237	PGSEMAAALATRLNRCVLAEPKLEAALAGVSPAGVICLWEAGAHEEAPAAQORVATEGL	4296
QY	56	-----	55
Db	4297	SVVQALDRDRAVRLMWMTMGAVAVEAGERVQVATAPVMGLGRTVMQERPELSCTLVLDPE	4356
QY	56	KMVLTEPEG :	65
Db	4357	ADAARSADVLLRELGRADDETQVAFRSGKRRVARLVKATTPGELLVPPDAESYRLEAQKG	4416
QY	66	-----	65
Db	4417	TLDQLRAPAQRAPGEGVEIKVTASGLNFRFTVLAVLGMVPGDAGMGGDCAGVATAVG	4476
QY	66	-----	65
Db	4477	QGVHRVAVGDVNMVTLGTLHRFTVTDARLVVROPAGLTPAQAAATVPVAFLTAMLALHDLGN	4536

```

Qy 66 ----- 65
Db 4537 LRRGERVLIHAAAGGVGAAVQIARWIGAEVFATASPSKWAAVQAMGVPRTHIASRTLE 4596
Qy 66 ----- 65
Db 4597 PAETFRVOTGGRGVVVLNALAGEFVDASLSLSTGGRFLEMKGTDIRDRAAVAAAHPGV 4656
Qy 66 ----- 65
Db 4657 RYRVFDILELAPDRTREILERVVEGFAAGHLRALPVHAFATKAEAAFRMAQARHOGKV 4716
Qy 66 ----- 65
Db 4717 VLLPAPSAAPLAPGTGVTLLTGGLGALGHVARMLAQGVPHMVLTCRRGLDTPGAAKAVA 4776
Qy 66 ----- 65
Db 4777 EIEALGARVTIAASDVADRNALEAVLQAIPAEMPLOGVIAHAGALDDGVLEQTTDRFSR 4836
Qy 66 --APNIT----- 70
Db 4837 VLAPKVTGANLHETAGNDLAFVFLFSSMSGLLGSAGOSNYAANTFLDALAAHRAEG 4896
Qy 71 -----ANLSSL----- 77
Db 4897 LAAQSLAWGPFSDCGMAAGLSAALQARLARHGMGALSQAQGTALLQALARPETQLGAMS 4956
Qy 78 ----- 77
Db 4957 LDVRAASQASGAAPVPPWRALVRAEARHTAAGAAGALAAARLGALEARRADEVRKVQAE 5016
Qy 78 ----- 77
Db 5017 IARVLSWSAASAVPDRPLSLDLGLSLTAVELRNVLGQVGTATLPATLAFDHPPTVDALTR 5076
Qy 78 ----- 77
Db 5077 WLDDKVLAEPSSVSAKSSPOVALDEPIAIIIGICRPGGVADPESFWRLEEGSDAVV 5136
Qy 78 ----- 77
Db 5137 EVPHERWDIDAFYDPPDPVGRKMTTRFGGFLSDIDRDPAPFGISPREATTMDPQORLLL 5196
Qy 78 ----- 77
Db 5197 ETSNEAFERAGILPERLMGSDTGVEVGLFYQEYAAALAGIEAFDGYLGTGTASVSGRI 5256
Qy 78 ----- 77
Db 5257 SYVLGLKPSLTVDTACSSSLVAVHLAQAALRRGECVALAGGVALMLTPATTFVEFSRLR 5316
Qy 78 ----- 77
Db 5317 GLAPGRCKSFAAADGVGWSGECAMLLKPLRDAQRDGPILAVIRGTAVNQGRSGL 5376
Qy 78 ----- 77
Db 5377 TAPNGSSQOEVIIRALEOAGLAPADVSYVECHGTGTTLGDPIEQVALCAVLAQGRPSDRP 5436
Qy 78 ----- 77
Db 5437 LVIGSVKSNIGHTQAAAGVAGVIKVALALERGLIPRSLHFDAPNPHIPWSELAVQVAAKP 5496
Qy 78 ----- 77
Db 5497 VENTRNGVPRAGVSSFGVSGTNAHVLEAPAAAFAPAAARSALFVLSAKSAALDAQ 5556
Qy 78 ----- 77
Db 5557 AARLSAHVVAHPELGLDLAFSLATRSPTYRLAVALSREALSALDAAQOQAPPA 5616
Qy 78 -----LSVC-----GWS----- 84

```

```

Db 5617 ARGHASTGSAPKVVVFVFFGOGSOMLGMQKLLSEEPVFRDALSACDRAIOAEGWLLAE 5676
Qy 85 ----- 84
Db 5677 LAADETTSQLGRIDVVQPALFAIEVALSALWRSWGVPEPDAVVGHSMGEVAHAHVAGALS 5736
Qy 85 ----- 84
Db 5737 EDVAIIICRRSLRLRRISGGEMAVVELSLAEAEALLGYEDRLSVSVNSPRSTVLAGE 5796
Qy 85 ----- 84
Db 5797 PAALAEVLALAAKGVFCRRVKVDVASHSPOIDPLRDELLAALGELEPROATVSMRSTVT 5856
Qy 85 ----- 84
Db 5857 STIMAGPELVASYWADNVROPVRAEAVOSLMEDGHCLFVEMSPHPLITTSVEIRRAK 5916
Qy 85 ----- 84
Db 5917 REGVAVGSLRRGODERLSMLEALGALWVGQAVGWERLFSAGGAGLRVRPLTPYQORER 5976
Qy 85 -----OTINPE----- 90
Db 5977 YWVDAPTGGAAGSRFAHAGSHPLLGEMOTLSTQSTRVWETTLDKRLPWLGDHRVOGA 6036
Qy 91 -----DDTD----- 94
Db 6037 VVFPAAYLEMALSSGAELGDGPLOVSDVLAELAFADDDTPAAVOVMATBERPORLOF 6096
Qy 95 -----PGH----- 97
Db 6097 HVASRVPGHGGAARFRSHARGVLRQIERAEVPARLDLAALRARLQASAPAAATYAALAE 6156
Qy 98 ----- 97
Db 6157 LEYPAFOGLVELWRGEALGRVLRPEAGSPAACRLHPALLDACFHVSSAFADRGEAT 6216
Qy 98 ----- 97
Db 6217 PMVPEIGSLRMFORPSELWCHARSVSHGKTPDRRSTDFWVVDSTGAIVAEISGLVAQ 6276
Qy 98 ----- 97
Db 6277 RLAGGVRRREDDWMEPAWEPTAVPGSEVMAGRMLLIGSGGLGAALHSALTAEAGHSV 6336
Qy 98 -----AD 99
Db 6337 HATGRGTSAAQLALLTASFQQAPTSVVHLGSLDERGVLDADAFDADALPESLVRGCD 6396
Qy 100 LVLY-----ITR----- 106
Db 6397 SVLMTVOAVAGAGFRDPPRLMLVTRGAOAIAGADVSAQAAPLLGLGLRVIALEHABLCAR 6456
Qy 107 ----- 106
Db 6457 IDLDPARRDEVDLELLAEALLADDAEEVAPRGGERRVARLVRRLPETDCREKIEPAERGP 6516
Qy 107 FOLEL----- 111
Db 6517 FRLEIDGSGVLDLVLVLRATERPPGPEVEIAVEAAGLNFUDVMRAMGIYPGPGDGPVAL 6576
Qy 112 ----- 111
Db 6577 GAECSGRIVAMGEVGSURIGQDVAVAPFSFGTHVTIDARMLAPRPAALTAQAALAPV 6636
Qy 112 ----- 111
Db 6637 AFMTAWYGLVHLRLRAGERVLIHSATGGTGLAAVQIARHLGAEIIPATAGTPEKRAWLRE 6696
Qy 112 -----PDGN----- 115

```

Db 6697 QGIAHVWDSRLDFAEQVLAATKGEQVDVVLNSLSGAAIDASLSTLVPDGRFIELGKTDI 6756
Qy 116 ----- 115
Db 6757 YADRSIGLAHFRKLSYSYSAVDLAGLAVRRPERVAALLAEVVDLLARGALQPLPVEIPPLS 6816
Qy 116 ----- 115
Db 6817 RAADAFRKAQAQHLGKLVLAELEDPVIRVPGESGVAIRADGAYLVTGGGLGSLSVAG 6876
Qy 116 ----- 118
Db 6877 WLAEOGAGHLVLVGRGAVSAEQQTAVAALAEHAGARVTVARADVADRAQMERILREVTAS 6936
Qy 119 ----- 126
Db 6937 GMPLRGVVHAAGILDGDLMOQTAPARFRAVNAKPVKGALHLHALTREAPLSFFVLYASGA 6996
Qy 127 ----- 135
Db 6997 GLLGSPGQGNVAAANTFLDALAHHRAOGLPALSDMGLFADVGLAAGQONRGARLVTRG 7056
Qy 136 ----- 143
Db 7057 TRSLTPDEGLWALERLLDGRTOAGVMPFDVQWVEFYPAASRRSLRLMTARRVASGR 7116
Qy 144 ----- 143
Db 7117 LAGDRDLRLATAEAGARAGMLOEVVRAQVSVLRLSEGKLDVDAPLTSIGMDSLMGLE 7176
Qy 144 ----- 148
Db 7177 LRNRIEAVLGITM 7189

RESULT 12

US-09-567-899-5
; Sequence 5, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-899-5

Query Match 27.0%; Score 209; DB 4; Length 7257;
Best Local Similarity 1.5%; Pred. No. 3.1;
Matches 93; Conservative 23; Mismatches 32; Indels 6165; Gaps 30;
Qy 1 AAGG----- 4
|||
Db 877 AAGGSVWPGVFTAGRVPLPTYPMQORQYWIIEPAEGLGATAADALAQMRYRVDWPEM 936
Qy 5 ----- 7
||| ILH-
Db 937 PRSSVDSRRARSGMLVLDRGVGEAAAAALSSQCCSCLVLAHAEASVAEQVTOALG 996
Qy 8 ----- 7

Db 997 GRNDWQGVLYLWGLDAVVEAGASABEVAKVTHLAAAPVLALQALGTGPRSPRLWIVTRG 1056
Qy 8 ----- 7
Db 1057 ACTVGGEPDAA PQAAALWGMGRVAALHHPGSGGLVDLPBESPTVEALVAELLSPOAE 1116
Qy 8 ----- 7
Db 1117 DQLAFQGRRRRAARLVAAPPEGNAAPVSLSAEGSYLVTGGLGALGLLVARMLVERGAGHL 1176
Qy 8 ----- 15
||| LLVAVG
Db 1177 VLISRHGLPDREWGRDQPPPEVRARIAAEIAEAQAGARVTVAAVDVDAEGMAALLAAVE 1236
Qy 16 P-----DVFOAHQ----- 23
|||
Db 1237 PPLRGVHAAGLLDDGLLAHQDAGRLARVLRPKVEGANVLHTLTREQPLDLVLFSSASG 1296
Qy 24 ----- 23
Db 1297 VFGSIGGSYAAGNAFLDALADLRLTQGLAALSTAWGLWAEGGMSOORREHEASGTWA 1356
Qy 24 ----- 23
Db 1357 MPTSRALAAMEWLLGTRATORVWIQMDWAHAGAAPROASRGRFWDRLVTATKEASSAVP 1416
Qy 24 ----- 23
Db 1417 AVERWRNASVETRSALVELVRGVVAGVMGFTDQGLDVRRGFAEOGLSLMAVEIRKRL 1476
Qy 24 ----- 23
Db 1477 QGELCMPLSATLAFDHTPTVERLVEYLLSQALEQDRTDVRSVRLPATEDPIAIVGAACRF 1536
Qy 24 ----- 29
||| EDTERY
Db 1537 PGVEDLESYMWQLLTEGVVSTEVPADRMNGADGRVPGSGEAGQRTVVPVPGFLREVET 1596
Qy 30 ----- 29
Db 1597 DAFFHISPREAMSLDPQORLLLEVSWEATERAGQDPSALRESPTGVFVGAGPNEYAERV 1656
Qy 30 ----- 29
Db 1657 QELADEAAGLYSGTGNMLSVAAGRLSFFLGLHGPTLAVDTACSSSLVALHLGQCSLRGE 1716
Qy 30 ----- 29
Db 1717 CDQALVGGVNMILLSPKTFALLSRMHALSPEGRCCKTFSADADGYARAEGCAMVVLKRLSDA 1776
Qy 30 ----- 29
Db 1777 QRDRDPILAVIRGTAINHDGPSSGLTVPSGPAQELLROALAHAGVVPADVDVECHGTG 1836
Qy 30 ----- 29
Db 1837 TALGDPIEVRALSDVYGOARPADRPILGAAKANLGHMEPAAGLAGLLKAVLALGOEQIP 1896
Qy 30 ----- 29
Db 1897 AQPELGELNPLLPWEALPVAVARAAPVWPRTDRPRFAGVSFGMSGTNAHVVLEAPAVE 1956
Qy 30 ----- 29
Db 1957 LWPAAPERSAELLIVLSGKSEGALDAQARLREHLDHWPBGLGDVAFSLATTRSANNHRL 2016
Qy 30 ----- 29
Db 2017 AVAVTSREGLLAALSAVAAGQTPPGAARCIASSSRGKLAFLTGOGAQTGPMGRGLCAAW 2076
Qy 30 ----- 29


```
Db 2077 PAFREAPDRCVALLFDRELDRLREVWMAEPGSAESLLLDQTAFTQPAFTVEYALTALWR 2136
QY 30 ----- 29
Db 2137 SWGVEPELVAGHSAGELVAACVAGVFSLEDGIVLVAARGRLMOGLSAGGAMVSLGAPEAE 2196
QY 30 ----- 29
Db 2197 VAAAVAPHAASVIAAVNGPEQVVIAGVEQAVQAIAGFAARGARTKRLHVSASHSPLM 2256
QY 30 ----- VLTNLN ----- 35
Db 2257 EPMLEBFRVAASVTYRRFSVLSVNLGKVADELSAGPYWVRHVREAVRFADGVKALH 2316
QY 36 ----- 35
Db 2317 EAGAGTFVEVGPKPTLLGLLPACLPPEAEPTLLASLPAGREEAAGVLEALGRLLWAAGGSVS 2376
QY 36 ----- 35
Db 2377 WFCVFTAGRRVLPPTYPMQORQYWPDIPEPSRRHAAADPTQGWFYRVDWPEIPRSLOKS 2436
QY 36 ----- 35
Db 2437 EBASRGSWLVLADKCGVGEAAVAAALSTRGLPCVVLHAPAEATSATAELVTEAAGRSRWQV 2496
QY 36 ----- IGAE ----- 39
Db 2497 VLYLWGLDAVWGAESAIDEIGDTRRATAPVLGARFLSTVSCSPRLWVTTRGACIVGDE 2556
QY 40 ----- 39
Db 2557 PAIAPCOAALWGMGRVAALHPGAWGGLVDLPDPRASPQASPIDGEMLVTELLSQETEDQ 2616
QY 40 ----- 39
Db 2617 LAFRHRRRHAARLVAAPOGQAAPVLSNEASLYVTGGLGLLIVAOWLVELGARHLVL 2676
QY 40 ----- 39
Db 2677 TSRRGLPDQAWCEQPPEIRARIAAVEALEARGARTVAAVDVADVEPMTALVSSVEPP 2736
QY 40 ----- 39
Db 2737 LRGVVHAAGVSMRPLAETDETLLLESVLURPKVAGSWLLHRLHGRPLDLFVLFSSGAHV 2796
QY 40 ----- 39
Db 2797 GSHSQAVAAANAFDGLAHLRRSOSLPALSVANGLWAEGMADAEAHARLSDIGVLPM 2856
QY 40 ----- 39
Db 2857 TSAALSALQRLVETGAAQRTVTRMDWARFAPVYTARGRRNLLSALVAGRDIIAPSPAAA 2916
QY 40 ----- 39
Db 2917 TRNRGLSVAEARVALHEIVHGAVARVLGFLDPSALDPGMCFNQGLDSLMAVEIRNLQ 2976
QY 40 ----- 39
Db 2977 AELDVRLSTTLAFDHTVQRLVEHLVDVLKLEDRSDTQHVRSIASDEPTAIVGAACRFP 3036
QY 40 ----- 39
Db 3037 GGVEDLESYVOLLAEVWVSAEVPADRWDAAWDYDPDPEIPGRYTVTKGAFRLDRLQDA 3096
QY 40 ----- 39
Db 3097 TFFRISPREANSLDPQORLLLEVSWEALESAGIAPDLRDSPTGVFVAGPNYYTORLR 3156
QY 40 ----- 39
Db 3157 GFTDGAAGLYCGTGNMLSVTAGRLSPFLGLHGTPLANDTACSSSLVALHLACOSLRIGEC 3216
QY 40 ----- 39
Db 3217 DOALVGVNVLLAPETTFVLLSRMALSPDGRCKTFSADADGARGEGCAVVVLKRLRDAO 3276
QY 40 ----- LLR ----- 42
Db 3277 RAGDSIILIRGSVNHDPGSSGLTVPNGPAOALLRQALSOAGVSPVDVDFVECHCTGT 3336
QY 43 ----- DP ----- 44
Db 3337 ALGDPIEVQALSEVYGPGRSDRPLVLGAANKANVAHLEAASGLASULLKAVLALRHEQIPA 3396
QY 45 ----- 44
Db 3397 QPELGEINPHLPWNTLPAVPRKAVPMGRGARPRRAGVSAGLSGTNVHVVLBEAPEVEP 3456
QY 45 ----- 44
Db 3457 APAAPRPVELVLSAKSAAALDAAAARLSAHLSAHPSELGLDVAFSLATTRSPMEHRLA 3516
QY 45 ----- 44
Db 3517 IATTSREALRGALDAAAQKTPQCAVRGKAVSSRGKLAFLFTGGAQMPGMRGLYETWP 3576
QY 45 ----- 44
Db 3577 AFREAPDRCVALPDREIDQPLREVWMAAPGLAQAAARLDOTAYAQPALFALEYALALWRS 3636
QY 45 ----- 44
Db 3637 WGVPHVLLGHSTIGELVAACVAGVFSLEDVAVRLVAARGRLMOALPAGGANVAIAASEAV 3696
QY 45 ----- SLGAQF ----- 50
Db 3697 AASVAPHAATVIAAVNGPDVAVIAGAEVQVLALGATFAARGIRTKRLAVSHAFHPLMD 3756
QY 51 ----- 50
Db 3757 PMLEDQFQVAATIAIRAPDRPVVSNVTGHVAGPEIATPEYVWRHVRSVAVRFGDGAKALHA 3816
QY 51 ----- 50
Db 3817 AGAATFVEVGPKPVLGLLPACLGEADAVLVPSLRADRSCEVVLALGAMVANGALDW 3876
QY 51 ----- RV ----- 52
Db 3877 KGVFPDGAARRVALPMYPMQREHMDLTSPSAAPAGIAGRWPAGVGLCMPGCVLHVLS 3936
QY 53 ----- HLV ----- 55
Db 3937 IGPRHQPFGLDHLVFGKVVVPGAFHVAVILSIAAERWPERAIELTGVFELKAIAEPDOE 3996
QY 56 ----- 55
Db 3997 VELHAVLTPEAAGDGLFELATLAAPETERRWTTHARGVQPTDGAPCALPRLEVLEDR 4056
QY 56 ----- 55
Db 4057 IQPLDFAGFLORLSAVRIGWGLWRWLODGRVGDSEASLATLVPTYNNAHDVAPLHPILLD 4116
QY 56 ----- 55
Db 4117 NGFAVSLLSRSEPDGTPPLPFAVERVWRWRAPVGRVRCGGVPRSQAFVCSFVLVDE 4176
QY 56 ----- 55
Db 4177 TGEVVAEVEGFCRRAPREVFLROESGASTAALYRLDWPAPLPDAPAEITEESWVVA 4236
QY 56 ----- 55
Db 4237 PGSEMAAALATRLNRCVLAEPKGLEAALAGVSPAGVTCLWEAGAHEEPAAORVATEGL 4296
```

```
QY 56 ----- 55
Db 4297 SVVQALRDRVRLMWVTMGAVAVEAGERVQVATAPWGLGRVTVMQERPELSCTLLVLEPE 4356
QY 56 ----- KMVILTEPEG ----- 65
Db 4357 ADAARSADVLLRELGRADDETQVAFRSGKRRVARLVKATTPEGLLVPDAESYRLEAGQKG 4416
QY 66 ----- 65
Db 4417 TLQDLRLAPQORRAPGPEVEIKVTASGLNFRVTLAVLGMYPDAGPMGGDCAGVATAVG 4476
QY 66 ----- 65
Db 4477 QGVRHVAVGDAVMTGLTLHRFVTVUDARLVVROPAGLTTPAQATVVPVAFLTAWLALHDLGN 4536
QY 66 ----- 65
Db 4537 LRRGERVLIHAAAGVGMAAVQIARWIGAEVFATSPSKMAAVQAMGVPRTHIASRTLE 4596
QY 66 ----- 65
Db 4597 FAETFRQVTGGRGVDVNLNLAGFVDASLSLLSTGGRFLEMGTDIRDRAAVAAAHPGV 4656
QY 66 ----- 65
Db 4657 RYRVFIIELAPDRTREILERVVEGFAGHLRALPVHAPAITKAEAAFRFMAQARHQKV 4716
QY 66 ----- 65
Db 4717 VLLPAPSAPLAPTGVLLTGGGALGLHVARLQAQGVPHVLTGRRGLDTFGAAKAVA 4776
QY 66 ----- 65
Db 4777 ETEALGARVTIAASVDARNALAVLQAIPEAWPLQGVIIHAAGALDDGVLDQTTDRFSR 4836
QY 66 --APNIT----- 70
Db 4837 VLAPKVTGANLHELTAGNDLAFVLFSSMSGLGSAGOSNYAAANTFLDALAAHRAEG 4896
QY 71 -----ANLTSSL----- 77
Db 4897 LAAQSLANGPWSGGGWAAGLSAALQARLARHGMGALSQAQGTALLGOALARPETQLGAMS 4956
QY 78 ----- 77
Db 4957 LDVRAASQAGAAVPPVWRALVRAEARHTAAGAQGALAAALGALPEARRADEVRKVQAE 5016
QY 78 ----- 77
Db 5017 IARVLSWSAASAVPDRPLSDGLDLSLTAVELRNVLGQRVGATLPATLAFDHPHTVDALTR 5076
QY 78 ----- 77
Db 5077 WLDDKVLVAEPSVSSAKSPQVALDEPIAIGICRFFGGVADPESFWRLLLEGSDAVV 5136
QY 78 ----- 77
Db 5137 EVPHERWDIDAFYDPDPDVRGKMTTRFGGFLSDIDRFDPAFFGISPREATTMDPOORLLL 5196
QY 78 ----- 77
Db 5197 ETSWEAFERAGILPERLMGSDTGVFVGLFYQEYAAALAGGIEAPDGLGTGTASVAGRI 5256
QY 78 ----- 77
Db 5257 SYVLGLKGPSLTVDTACSSSLVAVHLACQALRGECSVALAGGVALMLTPATFVFSRLR 5316
QY 78 ----- 77
Db 5317 GLAPDGRCKSFAAADGVGWSGECAMLLKPLRDAQRDGPILAVIRGTAVNQDGRSNGL 5376
QY 78 ----- 77

Db 5377 TAPNSSQQEVIRRALEQAGLAPADVSVECHGTGTTLGPQIEVQALGAVLAQGRPSDRP 5436
QY 78 ----- 77
Db 5437 LVIGSVKSNIGHTOAAAGVAGVIKVALALERGLIPRSLHFDAPNPHIPWSELAVOVAKP 5496
QY 78 ----- 77
Db 5497 VENTRNGVPRAGVSSFGVSGTNAHVVEEAPAAAAPAAAARSALFVLSAKSAAALDAQ 5556
QY 78 ----- 77
Db 5557 AARLSAHVVAHPGLGLDLAFSLATTRSPMTYRLAVAATSRALSAALDTAAOQOAPPA 5616
QY 78 -----LSVC-----GWS----- 84
Db 5617 ARGHASTGSAKVVVFPFGQGSQWLMGQKLLSEEPVFRDALSACDRAIQAEAGWSLLAE 5676
QY 85 ----- 84
Db 5677 LADETTSQLGRIDVVQPALFAIEVALSALWRSWGVPEPDAVVGHSMGVEAAAHVAGALSL 5736
QY 85 ----- 84
Db 5737 EDVAIIICRRSLLRRISGOGEMAVVELSLAEAEALLGYEDRLSVAVNSPRSTVLAGE 5796
QY 85 ----- 84
Db 5797 PAALAEVLAILAAGVFCRRVKVDVASHPSPOIDPLRDELLAALGELEPROQATVMSRSTV 5856
QY 85 ----- 84
Db 5857 STIMAGPELVASYADNVQRPVFAEAVQSLMEDGHGLFVEMSPHILTTSVBEIRRTK 5916
QY 85 ----- 84
Db 5917 REGVAVGSLRRGQDERLSMLEALGALVWVGQAVGWERLFSAGAGLRRVPLTYPWORER 5976
QY 85 -----OTINPE----- 90
Db 5977 YWVDAPTGAAGSRFAHAGSHPLLGEMOTLSTORSTRWETTLDLKLPLWGDHRVOGA 6036
QY 91 -----DDTD----- 94
Db 6037 VVFPQAYLENALSSGAELGDPLOVSDVVLAELAFADDTPAAVQVMATEERPQLQF 6096
QY 95 -----PGH----- 97
Db 6097 HVASRVPGHGGAAFRSHARGVLRQIERAEVPARLDLAALRARLOASAPAAATYAALAE 6156
QY 98 ----- 97
Db 6157 LEYGPFOGLVELWRGEALGRVRLPEAAGSPAACRUHPALLDACFHVSSAFADRGENT 6216
QY 98 ----- 97
Db 6217 PWPVETIGSLRWFORPSGELWCHARSVSHGKPTPDRRSTDFWVVDSTGCAIVAEISGLVAQ 6276
QY 98 ----- 97
Db 6277 RLAGVRRREBDDWMEPAWEPTAVPGSEVMAGRMLLICSGGGLGALHLSALTEAGHSV 6336
QY 98 -----AD 99
Db 6337 HATGRGTSAGLQALLTASFDGQAPTSVVHLGSLDERGVLDADAFDADALEESLVRGCD 6396
QY 100 LVLY-----ITR----- 106
Db 6397 SVLWTVQAVAGGFRDPPRLWLVTRGCAIAGAGDVSAQAQPLGLGLRVIALEHAELCAR 6456
QY 107 ----- 106
```

Db 6457 IDLPARRDGEVDELLAELLADDAEEVAFRGERRVARLVRLPETDCREKIEPAEGRP 6516
QY 107 FDLEL-----111
Db 6517 FRLEIDGSGVLDLVLRTERRPPGPGVEIAVEAAGLNFLDVMRAMGIYPGPGDGPVAL 6576
QY 112 -----111
Db 6577 CAECGRIVAMGESLRIGQDVVAVAPFSFGTHVTIDARMLAPRPAALTAQAALPV 6636
QY 112 -----111
Db 6637 AFMTAWGLVHLGRLAGERVLIHSATGGTGLAAVOIARHLGAEIFATAGTPEKRWLRE 6696
QY 112 -----115
Db 6697 QGIAHVWDSRLDFAEQVLAATKGEVDVVLNSLSGAIDAASLTLPDGRFELGKTDI 6756
QY 116 -----115
Db 6757 YADRSGLAHFRKSLSYSVADLAGLAVRRPERRVAALLAEVVDLLARGALQPLPVEIFPLS 6816
QY 116 -----115
Db 6817 RAADAFRMAQAQHLKGLVLALEDPDVIRVPGESVAIRADGAYLVVTGGLGLSVAG 6876
QY 116 -----118
Db 6877 WLAEOGAGHLVLCRSAGVSAEQQTAVAALEAHGARVTVARADVADRAQMERILREVTAS 6936
QY 119 ---RGVTQLGG-----126
Db 6937 GMPLRGVVAAGILDDGLLMQOTPARFRAVMKPVKRGALHLHALTREAPLSFFVLYASGA 6996
QY 127 -----ACSPWWSCL-----135
Db 6997 GLLGSPGQNYAAANTFLDALHRRRAQGLPALSIDMGLFADVGLAAGQNRGARLVTRG 7056
QY 136 ---ITEDTG-----143
Db 7057 TRSLTPDEGLWALERLLDGDRTQAGVMPFDVROWVEFPAAASSRRLSRLMTARRVASGR 7116
QY 144 -----143
Db 7117 LAGRDLLERLATAEAGARAGMLQEVVRAQVSVLRILSEGLDVPDAPLTLGLNDLSMLGLE 7176
QY 144 -----LQVTI 148
Db 7177 LRNRIEAVLGITM 7189

RESULT 13
US-08-460-751-2
; Sequence 2, Application US/08460751
; Patent No. 5891628
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen
; APPLICANT: Schneider, Michael
; APPLICANT: Gluckemann, Sandra
; TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
; TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,751
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/413,580
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7638-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4303 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-460-751-2

Query Match 26.9%; Score 208; DB 2; Length 4303;
Best Local Similarity 2.2%; Pred. No. 0.89;
Matches 87; Conservative 19; Mismatches 41; Indels 3862; Gaps 27;
QY 2 AGG-----1LH-----7
Db 23 AGGPRGCGPCBPCLCGPAPGACRVNCSGRGLRTLGLPALRIPADATELDVSNNLLRAL 82
QY 8 -----7
Db 83 DVGLLANLSALAEIDISNNKISTLEEGIFANLNLSEINLSGNPFECDCGLAWLPQWAE 142
QY 8 -----7
Db 143 QQVRVVQPEAATCAGPSLQAGPLLGIPLLDGCGEEYVACLPDNNSGTVAVVSAAHE 202
QY 8 -----7
Db 203 GLQPEACSAFCSTGGGLAALSEQWCLCGAQPSSAFCLSLCSGPPAPPAPTGRCP 262
QY 8 -----7
Db 263 TLLQHVFPASPGATLVGPHGLASQALAAFHIAPLVTDTRWDFGDSAEVDAAGPAAS 322
QY 8 -----LELL-----11
Db 323 HRYVLPGRYHVHTAVLALGAGSALLGTDVQVEAPAALELVCPSSVQSDLSLSTONRG 382
QY 12 -----11
Db 383 SGLEAAYSIVALGEBEPARAVHPLCPSPDTEIFPGNGCHCYRLVVEKAWLQAOEQCOAWAGA 442
QY 12 -----VAVGP-----16
Db 443 ALAMVDSPAVQRFVSRVTRSLDVMIGFSTVQGVVEGPAQGEAFSLSCQNWLPGEHP 502
QY 17 -----16
Db 503 ATAEHCVRGLGTGTCNTDLCSPHYSYCELOPGGPVQDAENLVGAPSGDLQGLPTPLAQ 562
QY 17 -DVFOAQHOEDTERYV-----LTNLNIGAEELRDP-----44
Db 563 QDGLSAPHEPVEVMVFPGLRLSREAFLLTAFFGTQELRRPAQLRLOVYVLLSTACTPENG 622
QY 45 -----44
Db 623 SEPESRSPDNRTOLAPACMPGGRWCMPGANICLPDASCHPOACANGCTSGPLPGAPVAL 682
QY 45 -----44

```
Db 683 WREFLFSVPAGPPAQYVTLHGQDVLMLPGDLVGLQHDAGPGALLHCSAPGHPGRAPY 742
QY 45 ----- 44
Db 743 LSNASSMLPHLPAQLEGTMGCPACALRLLAQREQTLVLLGLRPNPGLRLPCRYEVAEV 802
QY 45 ----- 44
Db 803 GNGVSRHNLSCSFDVVSPVAGLRVYPAPRDGRLVYPTNGSALVLQVDSGANATATARP 862
QY 45 --SLGAQR-----VHLVKMV----- 58
Db 863 GGSLSARFENVCALVATVPACPWETNDTLFSVVALPMLSGEHHVDDVVVENSASRAWL 922
QY 59 ----- 58
Db 923 SLRVTAEEPICGLRATPSPPEARVLQGLVRYSPVVEAGSDMVFRWTINDKQSLTFQNVVF 982
QY 59 ----- 58
Db 983 NVIYOSAAVKFLSLTASNHSNVTVNVTVERMNRMOGLOVSTVPAVLSPNATLALTAG 1042
QY 59 ----- 58
Db 1043 VLVDSAVEAFLWTFGDGEQALHQFPYPINESFPVDPDPSVAQVLVEHNVHTHYAAPGEVL 1102
QY 59 ----- 58
Db 1103 LTVLASAFENLTQQVPVSVRASLPSVAVGVSGLVAGRPVTFYPHPLPSGGVLYTWD 1162
QY 59 ----- 58
Db 1163 FGDGSPVLTSQPAANHITYASRGTYHVRLEVNNTVSGAAQAQDVRVFEELRGLSDMSLA 1222
QY 59 ----- 64
Db 1223 VEQGAPVVVSAAVQTGDNITWTGMDGTVLGSGPEATVEHVYLRQNCVTTVGAGSPAGH 1282
QY 65 -----GA 66
Db 1283 LARSLHLVFLVLELVRPEAACIPTQPDARLTAYVTGNPAHYLFDWTFDGSNTTVRC 1342
QY 67 PNITANLTSS-----LLSVC----- 81
Db 1343 PTVTHNFRSGTFPLALVLSRVNRHAHYFTSICVEPEVGNVTLOPERQFVLGDEAWLVA 1402
QY 82 ----- 81
Db 1403 CAMPPPPYRYTWDGTEBAAPTRARGPEVTFIYRDPGSYLVTVTASNNIISAANDSALVEV 1462
QY 82 ----- 81
Db 1463 QBPVLVTSIKVNGSLGLELQQVPLPSAVGRGPASVYLDLGDGWLGEPEVTHAYNSTGD 1522
QY 82 ----GW-----SQT----- 87
Db 1523 FTVRVAGMNEVRSEAWLNVTVKRRVRLGVNANSTRVPLNGSVFSTSLEAGSDVRYSW 1582
QY 88 ----- 87
Db 1583 VLCDRCTPIPGGPTISYTFRSVGTENIIVTAENEVGAQDSIFVYVVLQIEGLQVVGGOR 1642
QY 88 ----- 87
Db 1643 YFPTNHTVQLQAVRDGTNVYSWTAWDRGPALAGSGKGFSLTVLEAGTYHVQLRATNM 1702
QY 88 ----- 87
Db 1703 LGSAMADCTMDFEVPVGMWMAASPAAVNTSVTLSELAGSGGVYVYTWSEGLSWET 1762
QY 88 ----- 87
Db 1763 SEPTTHSFPPTPGLHLVTMTAGNPLGNSANATVEVDVQVPVSGLSIRASEPGGSFVAAGSS 1822
QY 88 ----- 87
Db 1823 VPFWQLATGTVNSWCWAVPGSSKRGPHVTMVPFDAGTFSIRLNASNAVSVSATYNLT 1882
QY 88 -----NPE----- 90
Db 1883 AEEPVLGLVLWASSKVAVPQOLVHFQILLAAGSAVTFRLOVGGANPEVLPGRFRSHSPR 1942
QY 91 ----- 90
Db 1943 VGDHVSVRGKXHVHWAQAQVRIVVLEAVSGLQVPNCCEPGIATGTERNFARVORGSRV 2002
QY 91 ----- 90
Db 2003 AYAWYFSLQKVGDSLVLISGRDVTYTPVAAAGLLEIQVRAFNALGSENRTLVLVEQDAVQ 2062
QY 91 -----DDTD-----PG----- 96
Db 2063 YVALQSGPCFTNRSQAPEAATSPSPRRVAYHWDGDSGGQDTDEPRAEHSYLRPGDYRV 2122
QY 97 -----HADL-----VLYI 104
Db 2123 QVNASNLSVFFVAQATVTVQVLACREPEVDVVLPLQVLMRRSQRYNLEAHVDLRDCVTYQ 2182
QY 105 TR----- 106
Db 2183 TEYRWEVYRTASCORPGRPARVALPGVDVSRPRLVLPRLALPVGHYCFVVFVSGDTPLT 2242
QY 107 -----PDELPDG----- 114
Db 2243 QSIQANVTVAERLVPITIEGGSYRVMSDTRDLVLDGSESYDPNLEDGQDTPLSFHWACVA 2302
QY 115 -----NRQV----- 118
Db 2303 STOREAGCALNFGPRGSSTVTIPRERLAAGVEYTFSLTVWKAGRKEATQTVLIRSGR 2362
QY 119 ----- 118
Db 2363 VPIVSECVCKAQAVEVSRSSVYLEGRCNCSGSKRGKRWAAFTFSNKTVLDETTT 2422
QY 119 -----RGVTO-----LGAC----- 128
Db 2423 STGSAGMRLVLRGVLRDGEGYTTTLTVLGRSGEBECCASIRLSPNRPPLGSCRLPLG 2482
QY 129 ----- 128
Db 2483 AVHALTTKVHPECTGWHDAEDAGAPLVYALLRRRCQGHCEFCVYKGLSSSYGAVLPPG 2542
QY 129 ----- 128
Db 2543 FRPHFVGLVAVVQDLGAAVVALNRSIAITLPEPNGSATGLTVMLHGLTASVLPGLLRQ 2602
QY 129 ----- 128
Db 2603 ADPOHVIEYSLAVTVLNEYERALDVAAABPKHQHRAQIRKNITETVLSLRVHTVDDIQ 2662
QY 129 -----SPT----- 131
Db 2663 QIAAALQCMGSPRELVCRCCLKQTLHKLEAMMLILOAETTAGTVTPTAIGDSILNITGD 2722
QY 132 ----- 131
Db 2723 LIHLASSDVRAPQPSSELGAESPRMVASOAYNLTSLMRILMRSRVLNEEPLTLAGEEIV 2782
QY 132 ----- 131
Db 2783 AOGKRSDRSLCYGGAPGPGCHFSIPEAFSGALANLSDVVQLFLVDSNPPFPFGYISNY 2842
QY 132 ----- 131
Db 2843 TVSTKVSMAFQTQGAQIPIERLASERAITVKVPNNSDWAARGHRSSANSANSVVVQPO 2902
```

```
QY 132 ----- 131
Db 2903 ASVCANVTLDSSNPAAGLHLQNLNTLLDGHYLSEEPYLAIVLHSEPRNEHNCASRR 2962
QY 132 -----WS----- 133
Db 2963 IRPESLOQADHRPVYTFISPGSRDPAGSYHLNLSHFRWSALQVSLYTLSCQYFSEED 3022
QY 134 ----- 133
Db 3023 MVRTEGLLPLEETSPRQAVCLTRHLTAFOGASLFVPPSHRVFVPEPTADVNVIVMLTCA 3082
QY 134 -CLIT----- 137
Db 3083 VCLVTYVMAAILHLKLDLDASRRAIPFCGQRGRFYKYLVTGHRGSGTTHVGMIL 3142
QY 138 ----- 137
Db 3143 YGVDSRSGRHLDDGDRAFHRNSLDIFRIATPHSLGVMKIRVMDNKGSLPAWFLQHVIV 3202
QY 138 ----- 137
Db 3203 RLOQTARSAFFLVNDWLSVETEANGGLVEKEVLAASDAALLRFRLLVAELQORGFDFKHI 3262
QY 138 ----- 137
Db 3263 WLSINDRPPRSRFRRIQATCCVLLICFLGANAVVYGVGDSAYSSTGHVSRSLPSLVDVT 3322
QY 138 ----- 137
Db 3323 VAVGLVSVVVPVYLAILFLFRMSRSKVAGSPPTPAQQVLDIDSCLDSSVLDSSFLT 3382
QY 138 ----- 137
Db 3383 FSLGHAQAFVQMKSDFLDPSKSLVCPWPSGEGTLWPDLLSDPSIVGSLNRLQARGQA 3442
QY 138 ----- 137
Db 3443 GHCLGPEEDGFSLASPYSPAKSFASDEDLIQOVLAEVSSPAPTQDTHMETDLSLSS 3502
QY 138 ----- 137
Db 3503 TPCEKTETLQRLGELGPPSPGLNWEQQAARLSRTGLVEGLRKLPAWCAASLHGLS 3562
QY 138 ----- 137
Db 3563 LLLVAVAVAGWVGASFPFGVSVAMLLSSSASFLASFLGWEPLKYLLEALYFSLVAKRL 3622
QY 138 --ED----- 139
Db 3623 HPDEDDLTVESPAVTPVSARVPRVRPPHGFALFLAKEEARKVRLHGMRLSLLYMLFLL 3682
QY 140 ----- 139
Db 3683 VTLASYGDSCHGHAYRLQSAIKQELHSRAFLAITRSEELWPMMAHVLLPYVHGNQSP 3742
QY 140 ----- 139
Db 3743 ELGPPRLROVRLQEALYDPDPPGRVHTCSAAGGFSTSDYVGWESPHNGSGTWAYSAPDL 3802
QY 140 ----- 139
Db 3803 LGANWGCANYDSGGYVQELGSLSESRDLRFLQLHNWLDNRSRAVLELTRYSPAVG 3862
QY 140 ----- 139
Db 3863 LHAAVTLRLEFPAGRALAALSVRPFALRRLSAGLSPLLTSTVCLLLFVAVHFAVAEARTW 3922
QY 140 -----TGPD----- 143
Db 3923 HREGRWRLRGAMARWLLVALTAATALVRLAQLGAADROWTRFVRGRPRRFTSPDOVAH 3982
```

```
QY 144 -----LGVTI 148
Db 3983 USSAARGLAASLLFLLLVKAAQHVRPVRQWSVFGKTLCLRALPELLGVTL 4031
RESULT 14
US-09-320-878-1
; Sequence 1, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320.878A
; EARLIER FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 1
; LENGTH: 4551
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-1
Query Match 26.8%; Score 207; DB 3; Length 4551;
Best Local Similarity 1.9%; Pred. No. 1.2;
Matches 84; Conservative 19; Mismatches 42; Indels 4181; Gaps 23;
QY 1 AAGG----- 4
Db 52 AAGGQAVTDVPADRWNAGDFYDPRDSAPGRSNRWGFIEDVDRFDAAFGISPREAEM 111
QY 5 ----- 4
Db 112 DPOORLALGLWEALERAGIDPSSLTGTRTCVAFAGAIWDDYATLKHROGGAITPHTVTG 171
QY 5 ----- 4
Db 172 LHRGIIANRLSYTLGLRPSMWVDSSGSSSLVAHVLAACESLRRGESELALAGVSLNLP 231
QY 5 ----- 4
Db 232 DSIIGASKFGLSPDGRAYTFDARANGYVRGEGGFVLRKLSRAVADGDPVLAVIRGSA 291
QY 5 ----- 4
Db 292 VNNGAAQGMTPDAQOAEAVLREAHERRAGTAPADVRYVELHGTGTVPDPIEAALGAA 351
QY 5 -----ILHLE----- 9
Db 352 LGTRPAGQPLLGVSVKTNIGHLEGAAGIAGLIKAVLAVRGRALPASLNYETPNPAIPFE 411
QY 10 ----- 9
Db 412 ELNLRVNTEYLPWEPEHGDGMVMVGVSSFGMGGTNAHVLEAPGVVEGASVVESTVGS 471
QY 10 ----- 9
```

```
Db 472 AVGGVPMVVSAAALDAQIERLAAFAASDRDTDGDAGAVDAGAVARVLGG 531
Qy 10 -----LLVAVGPD----- 17
Db 532 RAQFEHRAVVGSGPDDLAAALAAPEGLVRGVASGVGRVAFVFPQGTQWAGGAELDS 591
Qy 18 --VFOA----- 21
Db 592 SAVFAAACEAALSPLYVDWSLEAVVROAPGAPTLERVDVQVPTFAVMVSLARVQHH 651
Qy 22 ----- 21
Db 652 GVTPOAVGHSQGEIAAAVAGALSDDAARVVTLRKSIIAHLAKGKGMLSALSSEDV 711
Qy 22 ----- 21
Db 712 LERLAGFDGLSVAANGPTATVVGDPVQIEELARACEADGVRARVIPDYASHSQVEI 771
Qy 22 ----- 21
Db 772 IESELAEVLAGLSPOAPRVPPFFSTLEGAWITEPVLDDGGYWRNLRHVRGFAPAVETLATD 831
Qy 22 ----- 21
Db 832 EGFTHFEVSAHPVLTMLPGTVTGLATLRDNGGODRLVASLAEWANGLAVDWSPLP 891
Qy 22 ---HOED-----TER----- 28
Db 892 SATGHSDLPTVAFOTERHWLGEIEALAPAGEPAVQPAVLRTAAAPAELOLQVRIL 951
Qy 29 ----- 28
Db 952 DKVRAQTAQVLGYATGGQIEVDRTFREAGCTSLTGVDLRNRINAAGFVRMAPSMIFDPPT 1011
Qy 29 ----- 28
Db 1012 PEALAEQLLVVHGEAANPAGAPAPVAAAGAVDEPVAIVGMACRLPGGVASPEDLWRL 1071
Qy 29 ----- 28
Db 1072 VAGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYVRQGGFIENVAGFDAFFGISPREAL 1131
Qy 29 -----VVL 32
Db 1132 AMDPOORULLLETWEAVEADAGIDPTSLRGRQGVGTGANTHEYGPSLRDGGEGLOGYLLT 1191
Qy 33 N-----LNIGAEILR----- 42
Db 1192 GNTASVMSGRVSYTLGLEPALTVDTCSSSLVALHLAVQALRKGEVDMALAGGVAVMPT 1251
Qy 43 ----- 42
Db 1252 PQMFVEFSRQRLAGDGRSKAFASADGTSWSEGVGVLLVERLSDARRNGHQVLAVRGS 1311
Qy 43 ----- 42
Db 1312 AVNQDASNGLTAPNGPSQQRVIRRALADARLTTSDDVDVVEAHGTGRLGDPPIEAQALIA 1371
Qy 43 ----- 42
Db 1372 TYGQRDDEQPLRLGSLKSNIGHTQAAAAGVSGVIKWQAMRHGLLPKTLHVDEPSDQIDW 1431
Qy 43 -----DPSLG- 47
Db 1432 SAGAVELLTEAVDWPEKODGGLRAAVSSFGISGTAHVVLEAPVTVVEGASVVVEPSVG 1491
Qy 48 -----AQF 50
Db 1492 SAVGGVTPVWVSAAALDAQIERLAAFAASDRDTDDADAGAVDAGAVAHVLADGRAQF 1551
Qy 51 ----- 50
Db 1552 EHRAVALGAGADDLVQALADPDGLIRGTASGVGRVAFVFPQGTQWAGGAELLDSSAVF 1611
```

```
Qy 51 -----RVHLV----- 55
Db 1612 AAAMAECEAALSPLYVDWSLEAVVROAPGAPTLERVDVQVPTFAVMVSLARVQHHGVT 1671
Qy 56 ----- 55
Db 1672 QAVVGHSGEIAAAVAGALPLDDAARVVTLRKSIIAHLAKGKGMLSALNEDAVLERL 1731
Qy 56 ----- 55
Db 1732 SDFDGLSVAANGPTATVVGDPVQIEELAQACKADGFRARIIPVDYASHRSQVEIESE 1791
Qy 56 ----- 55
Db 1792 LAQVLAGLSPOAPRVPPFFSTLEGTWITEPVLDDGTWYRNLRHVRGFAPAIETLAVDEGFT 1851
Qy 56 ----- 55
Db 1852 HFVEVSAHPVLTMLPETVTGLTLRREOGGOERLVTSLAEAWNGLIPVAMTSLPATAS 1911
Qy 56 ----- 55
Db 1912 RPGLPTYAFOAERYWLENTPAALATGDDWRYRIDWKELPAEGERSTGLSGRWLAVTPED 1971
Qy 56 ----- 55
Db 1972 HSAQAAVLTALVDAGAKVEVLTAGADDDREALAARLTALTGTGFTGVVSLLDGLVPQV 2031
Qy 56 -----KMWILTEPE----- 64
Db 2032 AMVOALGDAGIKAPLMSVTOGAVSVGRGLDTPADPDRAMLGLGRVVALEHPERWAGLVLD 2091
Qy 65 ----- 64
Db 2092 PAQPDAAALAHVLTALSGATGEDQIAIRTTGLHARRLARAPLHGRRTDRMQPHGTVLIT 2151
Qy 65 -----GAPNITANLTSS----- 76
Db 2152 GGTGALGSHAARMAWHGAEBHLLVSRSGEOAPGATQLTAEALTASGARVITAAACDVADPH 2211
Qy 77 ----- 76
Db 2212 AMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTASOVRRRAHRAKAVGASVLDLTRDLD 2271
Qy 77 ----- 76
Db 2272 LDAFVLFSSVSTLGIPOGNYAPHNAYLDALAARRRATCRSAVSANGWPGWGGMAAGD 2331
Qy 77 -----LLSVCGWSQTI----- 87
Db 2332 GVAERLRNHGVPQMDPELAALESALGRDETAITVADIDWDRFYLAYSSGRPQPLVEEL 2391
Qy 88 ----- 87
Db 2392 PEVRIIDARDSATSCGGSSAOGANPLAERLAAAAAPGERTELLGLVRAQAAVLRMRS 2451
Qy 88 ----- 87
Db 2452 PEDVAADRAFKDIGFDSLACVAGVLRNRLTRATGLQLPATLVFDHPTPLALVSLRSBFLGD 2511
Qy 88 -----NPED----- 91
Db 2512 EETADARRSAALPATVAGAGAGAGTDDDDPIAIVAMSCRYPGDIRSPEDLWMLSEGG 2571
Qy 92 -----DTP----- 95
Db 2572 EGITPTPTDRGWDLDGLYDADPDALGRAYVREGGFLHDAAEFDAEPFGVSPREALMDPO 2631
Qy 96 ----- 95
Db 2632 QRMLLTTSWEAFERAGIEPASLRSGSTGVFGLSYDYAARVAPNAPRQVGEVLLTGSTPS 2691
```

```
QY 96 ----- 95
Db 2692 VASGRIAYTFLEGPATTVDTCSSILTAHLAVRALRSGECTMALAGGVAMMATPHFV 2751
QY 96 ----- 95
Db 2752 EFSQRALAPGRSKAFSADADGFAAEGVGLLVERLSDARRNGHPVLAVVGTAVNQD 2811
QY 96 ----- 95
Db 2812 GASNGLTAPNGPSOORVIRQALADARLAPGDI DAVETHGTGTSIGDPIEAQLOATYGKE 2871
QY 96 ----- 95
Db 2872 RPAERPLAIGSVKSNIGHTOAAAGAAGIIMVLAMRHGTLPKTLHADEPSPHVDWANSGL 2931
QY 96 ----- 95
Db 2932 ALVTEPIDWAGTGPRAAVSSFGISGTNAHVLEQAPDAAGEVLGADEVSEVSETVAMA 2991
QY 96 ----- 95
Db 2992 GTACTSEAESEASEAPAPGSRASLPGHLPHVLSAKDEQSLRGQAAALHAWLSEPA 3051
QY 96 ----- 95
Db 3052 DLSADGPARLDVGYTLATSRFAFAHRAAVTAADRDGFLDGLATLAQGGTSAHVHLDTA 3111
QY 96 ----- 102
Db 3112 RDGTAFLEFTGQSGRPGAGRELYDRHPVFARALDEICAHLDGHLLEPLLDVNFPAER 3171
QY 103 ----- 102
Db 3172 EAALLDETRYTOCALFALEVALFRLVESWGRPAALLGHSVGETAAAHVAGVFSLADAAR 3231
QY 103 ----- 108
Db 3232 LVAARGRLMOELPAGGAMLAQAAEDEIRVWLETEERYAGRLDVAAVNGPEAAVLSGDAD 3291
QY 109 ----- 108
Db 3292 AAREAEAYWSGLRRRTALRVSHAFHSAHMDGMLDGPRAVLETFEPRRSLTVVSNVTGL 3351
QY 109 ----- 108
Db 3352 AAGPDDLCDEYVWRHVHRTVRFIDGYRVLRLGVRTCLLELPGDGVLTANAADGLADTPA 3411
QY 109 ----- 108
Db 3412 DSAAGSPVGPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAHGTGPDWH 3471
QY 109 ----- 113
Db 3472 ANFAGSAHRVDLPTVSFRDRYWLDAADTAVDTAGLGLGTADHPHLLGAVVSLEPDRDG 3531
QY 114 ----- 124
Db 3532 LLLTGRSLRTHPLWADHAVLGSVLLPGAAVVELAAHAAESAGLRDVRLETLLEPLVLP 3591
QY 125 -GGA- ----- 134
Db 3592 HGGVELRVTVGAPAGEPGESAGDGPVSLHSRLADAPAGTAWSCHATGLLATDRPELP 3651
QY 135 ----- 134
Db 3652 VAPDRAAMWPPQAAEEVPLDGLYERLDGNGLAFLPFLQGLNAVVRVEGEVFADIALPAT 3711
QY 135 ----- 134
Db 3712 NATAPATANGGSAAPYGIHPALLDASLHATAVGLGVDEPELVRVPFWHSVTVHAAG 3771
QY 135 ----- 134
```

```
Db 3772 AAAARVRLASAGTDAVSLSLTDGGRPLVSVRLTLRPTVTDQAAASRVGGLMHRVAMRP 3831
QY 135 ----- 134
Db 3832 YALASSGEODPHATSYGPTAVLGKDELKVAALLESAGVEVGLYPLDLAALSODVAAGAPAP 3891
QY 135 ----- 134
Db 3892 RTVLAPLPAGPADGGAEGVGTAVRTLELLOQWLADLHLAGTRLLLVTRCAVRDPESGA 3951
QY 135 ----- 134
Db 3952 DGGEDLSHAAAGLVRTAQTENPGRFGLDLADASSYRTLPSVLSDAGLRDEPOLALH 4011
QY 135 ----- 141
Db 4012 DGTIRLARLASVRPETGTAAPALAPEGTVLLTGGTGGGLGVARHVVGWVRRLLLVSR 4071
QY 142 ----- 141
Db 4072 RCTDAPGADLVHELEALGADVSAACDVADREALTAVLDAIPAHEPLTAVVHTAGVLSD 4131
QY 142 ----- 141
Db 4132 GTLPSTMTEDVEHVRPKVDAAFLDLDTSTPAYDLAAAFVWFSSAAAVFGGAGOGAYAAA 4191
QY 142 ----- 141
Db 4192 NATLDALAWRRRAAGLPAUSLWGLWAE'TSGMTGELGOADLRMRSRAGIGGISAEGIAL 4251
QY 142 ----- 141
Db 4252 LDAAALRDDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRRPASAASATTA 4311
QY 142 ----- 141
Db 4312 GTAGTGTADGAAETAATAVTLADRAATVDGPARQRLLEFVVGVEAEVLGHARGHIDAER 4371
QY 142 --FDLG 145
Db 4372 GFLLDLG 4377
```

RESULT 15

```
US-09-105-537-31
; Sequence 31, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 4613
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-31
```

Query Match 26.8%; Score 207; DB 4; Length 4613;
Best Local Similarity 1.9%; Pred. NO. 1.2;
Matches 84; Conservative 19; Mismatches 42; Indels 4181; Gaps 23;

QY 1 AAGG----- 4

Db 114 AAGGQAVTDVPADRMNAGDFYDPRSAQGRSNRMGGFIEDVDRFDAAFFGISPREAAEM 173

```
QY 5 ----- 4
Db 174 DPQRLALELGMALERAGIDPSSLTGRTRGVFAGAIWDDYATLKHROGAAITPHTVTG 233
QY 5 ----- 4
Db 234 LHRGITANRLSYTLGLRGSMVVDSCQSSSLVAHVHACESLRRGESELALAGVSLNLP 293
QY 5 ----- 4
Db 294 DSIIGASKFGLSPDCGRAVTFDARANGYVRGEGGFVWLKLSRAVADGDPVLAVIRGSA 353
QY 5 ----- 4
Db 354 VNNGGAAQGWTFDAAQAEVLRBEAHERAGTAPADVRYVELHGTGTPVGDPTEAAALGAA 413
QY 5 ----- 9
Db 414 LGTRPAGQPLLVGSKVTNIGHLEGAAGIAGLIKAVLAVGRALPASLNYETPNPAIPFE 473
QY 10 ----- 9
Db 474 ELNLRVNTYLPWEPEHDCQRMVGVSSFGMGGTNAHVLEEPGCCRCGASVVESTVGS 533
QY 10 ----- 9
Db 534 AVGGGVVPMVWSAKSAAALDAQIERLAFAASDRDRTDGVVDAGAVDAGAVARVLGG 593
QY 10 ----- 17
Db 594 RAQFEHRAVVVSGPDDLAALAAPEGLVRGVASGVRVAFVFPQGTQWAGMGAELDS 653
QY 18 --VFQA----- 21
Db 654 SAVFAAANAACEAALSPYVDWSLEAVVROAPGAPTLERVDVVQVPTFAMVSLARVMQH 713
QY 22 ----- 21
Db 714 GVTPOAVVCHSOGETAAYVAGALSDDMARVVTLRSKIAAHLAKGKGMLSLALSEDAV 773
QY 22 ----- 21
Db 774 LERLAGFDGLSVAAVNGPTATVVSGDPVQIEELARACEADGVRARVIVPDYASHSRQVEI 833
QY 22 ----- 21
Db 834 IESELAEVLAGLSPOAPRVPPFFSTLEGAWITEPVLDDGYWYRNLRHRYGFAPAVETLATD 893
QY 22 ----- 21
Db 894 EGFTHFVEVSAHPVLTMLPGTVTGLATLRRONGQDRLVASLAEAWANGLAVDWSPLP 953
QY 22 --HOED-----TER----- 28
Db 954 SATGHSDLPYAFQTERHWLGEIEALAPAGEPAVQPAVLRTAEAPAEALDRDEQLRVIL 1013
QY 29 ----- 28
Db 1014 DKVQRTAQVLGYATGGQIEVDRTFREAGCTSLTGVDLNRNINAAGFVRMAPSMIFDPPT 1073
QY 29 ----- 28
Db 1074 PEALAEQLLVHGEAANPAGAEPAVAAAGAVDEPAIVCMACRLPGGVASPEDLWRL 1133
QY 29 ----- 28
Db 1134 VAGGDAISEFPQDRGWDVEGLYHPDPEHPGTSYVROGFFIENVAGFDAFFGISPREAL 1193
QY 29 -----VYLT 32
Db 1194 AMDPOORLLETSWAEVADAGIDPTSLRGRQVGVGTGANTHEYGPSLRDGGEGDGLYLT 1253
QY 33 N-----LNIGAELLR----- 42
Db 1254 GNTASVMSGRVSYTLGLEPALTVDTACSSSLVALHVAQALRKRGVDMALAGGVAMPT 1313
QY 43 ----- 42
Db 1314 PCMFVFSRQRLAGDGRSKAFAASADGTSWSEGVGVLLVERLSDARRNGHOVLAVRGS 1373
QY 43 ----- 42
Db 1374 ALNQDGASNGLTAPNGPSQORVIRRALADARLTSTDVDVVEAHGTGRLGDPIEAQALIA 1433
QY 43 ----- 42
Db 1434 TYGGRDDEOPLRLGSLKSNIGHTQAAAGVSVTKMVQAMRHGLLTKTLHVDEPSDDIDW 1493
QY 43 -----DPSLG- 47
Db 1494 SAGAVELLTEAVDWPEKQDGLRRAAVSSFGISCTNAHVLEEAPVVVEGASVVEPSVG 1553
QY 48 -----AQF 50
Db 1554 SAVGGVTPMWVWSAKSAAALDAQIERLAFAASDRDRTDADAGAVDAGAVHVLADGRAQF 1613
QY 51 ----- 50
Db 1614 EHRAVALGAGADLLVQALADPDGLIRCTASGVGRVAFVFPQGTQWAGMGAELDSAVF 1673
QY 51 -----RVHLV----- 55
Db 1674 AAAMAECEAALSPPYDWSLEAVVROAPGAPTLERVDVVQVPTFAMVSLARVMQHGVTP 1733
QY 56 ----- 55
Db 1734 QAVVHSGEIAAAYVAGALPLDDAARVVTLRSKIAAHLAKGKGMLSLALNEDAVLERL 1793
QY 56 ----- 55
Db 1794 SDFDGLSVAAVNGPTATVVSGDPVQIEELAQACKADGFRARIIPVDYASHSRQVEIESE 1853
QY 56 ----- 55
Db 1854 LAQVLAGLSPOAPRVPPFFSTLEGTWITEPVLDDGYWYRNLRHRYGFAPAIETLAVDEGPT 1913
QY 56 ----- 55
Db 1914 HFVEVSAHPVLTWLTPEVTGLTLRREOGQERLVTSLABAWWNGLPVAVNTSLLPATAS 1973
QY 56 ----- 55
Db 1974 RPLPTYAFOERYWLENTPAALATGDDWRYRIDWKRLPAEGSERTGLSERWLAVTPE 2033
QY 56 ----- 55
Db 2034 HSAQAAAVLTALVDAGAKVEVLTAGADDDREALAARLTALTGDTGTVVSLDGLVPQV 2093
QY 56 -----KMWILTEPE----- 64
Db 2094 ANVQALGDAGIKAPLWSVTOGAVSVGRDLTPADPDRAMLWGLGRVVALEHPERWAGLVDL 2153
QY 65 ----- 64
Db 2154 PAQPDAAALAHVLTALSGATGEDQIAIRTTGLHARRLARAPLHGRPRTRDWPQHGTVLIIT 2213
QY 65 -----GAPNITANLTSS----- 76
Db 2214 GGTGALGSHAARMAWHGAHLLVSRSGEQAPGATQLTAELTASGARVTIAACDVADPH 2273
QY 77 ----- 76
Db 2274 AMRTLLDAIPAETPLTAVVHTAGALDDGIVDTLTAEQVRRRAHRAKAVGASVLDELTRDLD 2333
QY 77 ----- 76
```



```
Db 2334 LDAFVLSSVSTLIGPGQNYAPHNAYLDALARRRATGRSAVSAMGPDGGMAGD 2393
Qy 77 -----LLSVCGWSQTI----- 87
Db 2394 GVAERLNHGVPGMPELAALAESALGRDETAITVADIDWDRFYLAYSSGRPOPLVEEL 2453
Qy 88 ----- 87
Db 2454 PEVRIIDARDSATSGOGSSAQGANPLAERLAAAPGTERTEILLGLVRAQAAVLRMS 2513
Qy 88 ----- 87
Db 2514 PEDVAADRAFKDIGFDSLACVGLNRNLTRATGLQLPATLVDFDHPPTLALVLSLRSEFLGD 2573
Qy 88 -----NPED----- 91
Db 2574 EETADARRSAALPATVCGAGAGAGAGTADDDPIAIVAMSCRYPCDIRSPEDLWRMLSEGG 2633
Qy 92 -----DTDP----- 95
Db 2634 EGITPFTDRGWLGLYDADPDALGRAYVREGGFLHDAAEFDABEFGVSPREALAMDQ 2693
Qy 96 ----- 95
Db 2694 QRMLLTTSWEAFERAGIEPASLRGSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPS 2753
Qy 96 ----- 95
Db 2754 VASGRIAYTFGLEGPATTVDTCSSSLTALHLAVRALRSGECTMALAGGVAMATPHMFV 2813
Qy 96 ----- 95
Db 2814 EFSQRALAPDGRSKAFSADADFGAAEGVGLLVERLSDARRNGHPVLAVVRGTAVNQD 2873
Qy 96 ----- 95
Db 2874 GASNGLTAPNGSQORVIRQALADARLAPGDIIDAVETHGTGTSGLDPIEAQGLQATYKGE 2933
Qy 96 ----- 95
Db 2934 RPAERPLAIGSVKSNIGHTQAAAGAGIIMVLMRHGTLPKTLHADESPHVDWANSGL 2993
Qy 96 ----- 95
Db 2994 ALVTEPIDWPGTPRAAVSFGISGTNAHVLEQAPDAAGEVLGADEVPEVSETVAMA 3053
Qy 96 ----- 95
Db 3054 GTACTSEVAEGSEASEAPAPGSRASLPCHLPWVLSAKDEQSLRGQAAALHAWLSEPA 3113
Qy 96 ----- 95
Db 3114 DLSADGPARLRDVGTYLTATSRFAHRAAVTAADRDGFLDGLATLAQGGTSAHVHLDTA 3173
Qy 96 -----GHADIVL----- 102
Db 3174 RDGTTAFUFTGGSORPGAGRELYDRHPVFARALDEICAHLDGHLELPLLDVMAFAGESA 3233
Qy 103 ----- 102
Db 3234 EAALLDETRYTCALFALEVALFRLVESWGMRPALLGHSVGEINAAHVAGVFLADAAR 3293
Qy 103 -----YITRFD----- 108
Db 3294 LVAARGRLWQELPAGGAMLVAQAADEIRVMLETEERYAGRLDVAAVNGPEAAVLSDGD 3353
Qy 109 ----- 108
Db 3354 AAREAEYWSGLGRTRRALRVSHAFSAHMDGMLDGFRAVLETVEFRPRLTVVSNVTGL 3413
Qy 109 ----- 108
Db 3414 AAGPDDLCDPEYVWRHVRGTVRFILDGVRVLRDLGVRTCLGLPGDGLVTAMAADGLADTPA 3473
```

Search completed: March 20, 2003, 12:41:25
Job time : 149 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:37:58 ; Search time 18 Seconds
(without alignments)
439.578 Million cell updates/sec

Title: SEQID_1_15FUSED
Perfect score: 773
Sequence: 1 AAGGILHLELLVAAGPDVFQ.....SPTWSCLITDGTGDLGVIT 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.0

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 221153

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW PUB.pap:
2: /cgn2_6/ptodata/2/pubpaa/PCT NEW PUB.pap:
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW PUB.pap:
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap:
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW PUB.pap:
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap:
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap:
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap:
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW PUB.pap:
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pap:
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW PUB.pap:
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pap:
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW PUB.pap:
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	773	100.0	148 10 US-09-833-328-4 Sequence 4, Appli
2	773	100.0	150 10 US-09-833-328-6 Sequence 6, Appli
3	773	100.0	242 10 US-09-978-979-1 Sequence 1, Appli
4	773	100.0	1416 10 US-09-836-712-2 Sequence 2, Appli
5	770	99.6	203 10 US-09-781-0808-2 Sequence 2, Appli
6	770	99.6	1120 10 US-09-781-0808-11 Sequence 11, Appli
7	716	92.6	136 10 US-09-833-328-15 Sequence 15, Appli
8	702	90.8	133 10 US-09-833-328-2 Sequence 2, Appli
9	272	35.2	26926 9 US-09-759-508B-2 Sequence 7, Appli
10	237	30.7	997 10 US-09-918-171A-7 Sequence 6, Appli
11	235	30.4	11877 9 US-09-860-846-6 Sequence 6, Appli
12	235	30.4	11877 10 US-09-861-289-6 Sequence 6, Appli
13	231	29.9	1690 10 US-09-788-043C-5 Sequence 5, Appli
14	229	29.6	2150 10 US-09-321-987B-2 Sequence 2, Appli
15	229	29.6	2165 10 US-09-800-729-155 Sequence 155, App
16	228	29.5	7968 9 US-10-077-130-5 Sequence 5, Appli
17	222	28.7	730 9 US-10-163-547-16 Sequence 16, Appli
18	220	28.5	6304 9 US-10-147-026-16 Sequence 16, Appli
19	215	27.8	4349 9 US-10-160-758-15 Sequence 15, Appli

ALIGNMENTS

RESULT 1
US-09-833-328-4
; Sequence 4, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Schefflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Rndolf
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Gerritsen, Helena E.
; TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protea
; FILE REFERENCE: 247.00CIP
; CURRENT APPLICATION NUMBER: US/09/833,328
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254
; PRIOR FILING DATE: 2000-11-22
; SOFTWARE: Patent in version 3.1
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 4
; LENGTH: 148
; TYPE: PRT
; ORGANISM: human
US-09-833-328-4

Query Match 100.0%; Score 773; DB 10; Length 148;
Best Local Similarity 100.0%; Pred. No. 4.4e-31;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AAGGILHLELLVAAGPDVFQAHQEDTERVYLTNLNIGAEILLRDPISLGAQFRVHLVKMVL 60
DB 1 AAGGILHLELLVAAGPDVFQAHQEDTERVYLTNLNIGAEILLRDPISLGAQFRVHLVKMVL 60
OY 61 TEPEGAPNITANLTSLLSVCGWSQTIINPEDDTPDGHADLVLYITRFDLELPDGNQVRG 120
DB 61 TEPEGAPNITANLTSLLSVCGWSQTIINPEDDTPDGHADLVLYITRFDLELPDGNQVRG 120

```
QY 121 VTOLGGACSPWTSCLITDGTGDLGVTI 148
DB 121 VTOLGGACSPWTSCLITDGTGDLGVTI 148

RESULT 2
US-09-833-328-6
; Sequence 6, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Scheiflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Rndolf
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Gerritsen, Helena E.
; TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Acti
; TITLE OF INVENTION: Comprising a Polypeptide Chain with the Amino Acid Sequence AAGG
; FILE REFERENCE: 247.00CIP
; CURRENT APPLICATION NUMBER: US/09/833.328
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 150
; TYPE: PRT
; ORGANISM: human
; US-09-833-328-6

Query Match 100.0%; Score 773; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 4.6e-31;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLVAVGPDVFOAQHOEDTERVYLTNLNIGAEALLRDPDSLGAQFRVHLVKMVL 60
DB 3 AAGGILHLELLVAVGPDVFOAQHOEDTERVYLTNLNIGAEALLRDPDSLGAQFRVHLVKMVL 62

QY 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGDGNRVRG 120
DB 63 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGDGNRVRG 122

QY 121 VTOLGGACSPWTSCLITDGTGDLGVTI 148
DB 123 VTOLGGACSPWTSCLITDGTGDLGVTI 150

RESULT 3
US-09-978-979-1
; Sequence 1, Application US/09978979
; Patent No. US20020151702A1
; GENERAL INFORMATION:
; APPLICANT: Racie, Lisa, A.
; APPLICANT: Twine, Natalie, C.
; APPLICANT: Agostino, Michael, J.
; APPLICANT: Wolfman, Neil
; APPLICANT: Morris, Elisabeth
; TITLE OF INVENTION: Aggreganase Molecules
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Home Products Corporation
; STREET: One Campus Drive
; CITY: Parsippany
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

QY 121 VTOLGGACSPWTSCLITDGTGDLGVTI 148
DB 121 VTOLGGACSPWTSCLITDGTGDLGVTI 148

RESULT 4
US-09-836-712-2
; Sequence 2, Application US/09836712
; Patent No. US20010049106A1
; GENERAL INFORMATION:
; APPLICANT: PFIZER INC.
; TITLE OF INVENTION: ADAMTS POLYPEPTIDES, NUCLEIC ACIDS ENCODING THEM, AND
; FILE REFERENCE: PC10851A
; CURRENT APPLICATION NUMBER: US/09/836,712
; CURRENT FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1416
; TYPE: PRT
; ORGANISM: Human
; US-09-836-712-2

Query Match 100.0%; Score 773; DB 10; Length 1416;
Best Local Similarity 100.0%; Pred. No. 8.1e-29;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLVAVGPDVFOAQHOEDTERVYLTNLNIGAEALLRDPDSLGAQFRVHLVKMVL 60
DB 98 AAGGILHLELLVAVGPDVFOAQHOEDTERVYLTNLNIGAEALLRDPDSLGAQFRVHLVKMVL 157

QY 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGDGNRVRG 120
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/978,979
; FILING DATE: 16-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/60/241,469
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Kapinos, Ellen J.
; REGISTRATION NUMBER: 32,345
; REFERENCE/DOCKET NUMBER: GI 5435P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (973) 660-5000
; TELEFAX: (973) 683-4117
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 242 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
; US-09-978-979-1

Query Match 100.0%; Score 773; DB 10; Length 242;
Best Local Similarity 100.0%; Pred. No. 1.4e-30;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLVAVGPDVFOAQHOEDTERVYLTNLNIGAEALLRDPDSLGAQFRVHLVKMVL 60
DB 42 AAGGILHLELLVAVGPDVFOAQHOEDTERVYLTNLNIGAEALLRDPDSLGAQFRVHLVKMVL 101

QY 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGDGNRVRG 120
DB 102 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGDGNRVRG 161

QY 121 VTOLGGACSPWTSCLITDGTGDLGVTI 148
DB 162 VTOLGGACSPWTSCLITDGTGDLGVTI 189

RESULT 4
US-09-836-712-2
; Sequence 2, Application US/09836712
; Patent No. US20010049106A1
; GENERAL INFORMATION:
; APPLICANT: PFIZER INC.
; TITLE OF INVENTION: ADAMTS POLYPEPTIDES, NUCLEIC ACIDS ENCODING THEM, AND
; FILE REFERENCE: PC10851A
; CURRENT APPLICATION NUMBER: US/09/836,712
; CURRENT FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1416
; TYPE: PRT
; ORGANISM: Human
; US-09-836-712-2

Query Match 100.0%; Score 773; DB 10; Length 1416;
Best Local Similarity 100.0%; Pred. No. 8.1e-29;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLVAVGPDVFOAQHOEDTERVYLTNLNIGAEALLRDPDSLGAQFRVHLVKMVL 60
DB 98 AAGGILHLELLVAVGPDVFOAQHOEDTERVYLTNLNIGAEALLRDPDSLGAQFRVHLVKMVL 157

QY 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELPGDGNRVRG 120
```

Db 158 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELDPGNQVRG 217
Qy 121 VTOLGGACSPWTSCLITDGTGFDLGVTI 148
Db 218 VTOLGGACSPWTSCLITDGTGFDLGVTI 245

RESULT 5
US-09-781-080B-2
; Sequence 2, Application US/09781080B
; Patent No. US20020142439A1
; GENERAL INFORMATION:
; APPLICANT: HOLLOWAY, JIM
; APPLICANT: SHEPPARD, PAUL
; APPLICANT: YAMAMOTO, GAYLE
; TITLE OF INVENTION: Anti-Angiogenic Intestinal Peptides,
; TITLE OF INVENTION: zdn125
; FILE REFERENCE: 99-82
; CURRENT APPLICATION NUMBER: US/09/781,080B
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-781-080B-2

Query Match 99.6%; Score 770; DB 10; Length 203;
Best Local Similarity 99.3%; Pred. No. 1.3e-30;
Matches 147; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGGILHLELLVAVGPDVFOAHOEDTERVYLTNLNIGAEILLRDPSSLGAQFRVHLVKWVIL 60
Db 2 AAGGILHLELLVAVGPDVFOAHOEDTERVYLTNLNIGAEILLRDPSSLGAQFRVHLVKWVIL 61
Qy 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELDPGNQVRG 120
Db 62 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELDPGNQVRG 121
Qy 121 VTOLGGACSPWTSCLITDGTGFDLGVTI 148
Db 122 VTOLGGACSPWTSCLITDGTGFDLGVTI 149

RESULT 6
US-09-781-080B-11
; Sequence 11, Application US/09781080B
; Patent No. US20020142439A1
; GENERAL INFORMATION:
; APPLICANT: HOLLOWAY, JIM
; APPLICANT: SHEPPARD, PAUL
; APPLICANT: YAMAMOTO, GAYLE
; TITLE OF INVENTION: Anti-Angiogenic Intestinal Peptides,
; TITLE OF INVENTION: zdn125
; FILE REFERENCE: 99-82
; CURRENT APPLICATION NUMBER: US/09/781,080B
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 1120
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: (1)...(1120)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-781-080B-11

Query Match 99.6%; Score 770; DB 10; Length 1120;
Best Local Similarity 99.3%; Pred. No. 6.7e-29;
Matches 147; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGGILHLELLVAVGPDVFOAHOEDTERVYLTNLNIGAEILLRDPSSLGAQFRVHLVKWVIL 60
Db 105 AAGGILHLELLVAVGPDVFOAHOEDTERVYLTNLNIGAEILLRDPSSLGAQFRVHLVKWVIL 164
Qy 61 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELDPGNQVRG 120
Db 165 TEPEGAPNITANLTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELDPGNQVRG 224
Qy 121 VTOLGGACSPWTSCLITDGTGFDLGVTI 148
Db 225 VTOLGGACSPWTSCLITDGTGFDLGVTI 252

RESULT 7
US-09-833-328-15
; Sequence 15, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Scheifflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Randolph
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Gerritsen, Helena E.
; TITLE OF INVENTION: Compositing a Polypeptide Chain with the Amino Acid Sequence
; FILE REFERENCE: 247,00CIP
; CURRENT APPLICATION NUMBER: US/09/833,328
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 136
; TYPE: PRT
; ORGANISM: human
US-09-833-328-15

Query Match 92.6%; Score 716; DB 10; Length 136;
Best Local Similarity 100.0%; Pred. No. 2.4e-28;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 13 AVGPDVFOAHOEDTERVYLTNLNIGAEILLRDPSSLGAQFRVHLVKWVILTEPEGAPNITAN 72
Db 1 AVGPDVFOAHOEDTERVYLTNLNIGAEILLRDPSSLGAQFRVHLVKWVILTEPEGAPNITAN 60
Qy 73 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELDPGNQVRGVTOLGGACSPW 132
Db 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDLELDPGNQVRGVTOLGGACSPW 120
Qy 133 SCLITDGTGFDLGVTI 148
Db 121 SCLITDGTGFDLGVTI 136

RESULT 8
US-09-833-328-2
; Sequence 2, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Scheifflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Randolph
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus

APPLICANT: Furlan, Miha
APPLICANT: Turecek, Peter
TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
FILE REFERENCE: 247.00CIP
CURRENT APPLICATION NUMBER: US/09/833,328
CURRENT FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 09/721,254
PRIOR FILING DATE: 2000-11-22
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 133
TYPE: PRT
ORGANISM: human
US-09-833-328-2

Query Match 90.8%; Score 702; DB 10; Length 133;
Best Local Similarity 100.0%; Pred. No. 1.1e-27;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 PDVFOAQHEDTERVLTNLNIGALLRPSLGAQFRVHLVKMVLTEPEGAPNITANLTS 75
DB 1 PDVFOAQHEDTERVLTNLNIGALLRPSLGAQFRVHLVKMVLTEPEGAPNITANLTS 60
QY 76 SLLSVCWSQTNIPEDDTPGHADLVLYITRFDELDPGNROVRGVTOLGGACSPWSCL 135
DB 61 SLLSVCWSQTNIPEDDTPGHADLVLYITRFDELDPGNROVRGVTOLGGACSPWSCL 120
QY 136 ITEDTGFDLGVTI 148
DB 121 ITEDTGFDLGVTI 133

RESULT 9
US-09-759-508B-2
Sequence 2, Application US/09759508B
Publication No. US20020182599A1
GENERAL INFORMATION:
APPLICANT: Fishman, Mark C.
TITLE OF INVENTION: Methods for Diagnosing and Treating Heart Disease
FILE REFERENCE: 00786/381002
CURRENT APPLICATION NUMBER: US/09/759,508B
CURRENT FILING DATE: 2001-01-12
PRIOR APPLICATION NUMBER: US 60/175,787
PRIOR FILING DATE: 2000-01-12
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 26926
TYPE: PRT
ORGANISM: Homo sapiens
US-09-759-508B-2

Query Match 35.2%; Score 272; DB 9; Length 26926;
Best Local Similarity 0.4%; Pred. No. 0.4;
Matches 101; Conservative 16; Mismatches 23; Indels 24887; Gaps 28;
QY 9 ELL----- 11
DB 190 ELLVQGEVEEVPKTKTIVTAQISESROTRIEKKIEAHFARSIAIVEMVIDGAAGQOL 249
QY 12 ----- 11
DB 250 PHKTPPIPPKPSRPTPPSIAAKAQLARQOSPPIRHSFSPVRHVRAPTPSPVRSVP 309
QY 12 -----VAVGPDV----- 18
DB 310 AARISTSPVRSPLLMKRTQASTVATGPEVPPPWKQEGVASSSEAMRETTLTSTQ 369
QY 19 ----- 18

Db 370 IRTEERMEGRYGVQEQVTISGAAGAAASVSASAYAAEAVATGAKEVKODADKSAAVATV 429
QY 19 ----- 18
Db 430 VAAVDMARVREPVISAVEOTAORTTTTAVHIQPAQEOVRKEAKTAVTKVVAAADKAKEQ 489
QY 19 ----- 18
Db 490 ELKSRTKEIITTKQEQMHVTHEQIRKETKTFVPKVVISAAKAKEQETRISEITKKQKQ 549
QY 19 ----- 18
Db 550 VTOBAIMKETRTKTVVPKVIATPKVKEODLVSRGREGITTKRBOVOITOFKMRKEAKTA 609
QY 19 ----- 18
Db 610 LSTIAVATAKAKEQETILTRTETMATRBOQIOVTHGKVDVGKAEAVATVAAVDOARVR 669
QY 19 ----- 18
Db 670 EPREPGHLEESYAOQTTEYGYKERISAAKVAEPPORPASEPHVVPKAVKPRVIOAPSET 729
QY 19 ----- 18
Db 730 HIKTDOKGHMISSQIKKTTDLTTERLVHVKRPRTASPHFTVSKISVPKTEHGYEASIA 789
QY 19 ----- 18
Db 790 GSAIATLOKELSATSSAOKITKSVKAPTVPKSETRVRAEPTPLQPPFPADTPTDYKSEAG 849
QY 19 ----- 18
Db 850 VEYKEVGVSITGTTVREERFEVLHGREAKVTETARVPAPVEIPVTPPTLVSLKXNVTVI 909
QY 19 -----FO-----AQEDTERYV- 30
Db 910 EGESVTLECHISGYSPPTWYREDYQIESSIDFOITFOSGIARLMIREAPFDSGRFTC 969
QY 31 ----- 30
Db 970 SAVNEAGTVSTCYLAVQVSEBEFETKETTAVTEKFTTEBKRFRVESRDVMTDTSLTEQAG 1029
QY 31 ----- 30
Db 1030 PGEPAAPYFITKPVVQKLVVEGSGVFCQVGNPKPHVYKKSGLTGTGYRYKVSYNKQ 1089
QY 31 ----- 30
Db 1090 TGECKLVISMTFADDAGEYTVVVRNKHGETSASALLSEADYELLMKSOQEMLYQTVTA 1149
QY 31 ----- 30
Db 1150 FVQPEVGETAPGFVYSEYEKEQALIRKMAKDTVVVTVVEDOEPHISSEERLI 1209
QY 31 ----- 30
Db 1210 KEIFRYIKTTLLEELLEDGEKMAVDISEASESGDLRIKXRYILEGMVGFPHCKWS 1269
QY 31 ----- 30
Db 1270 GYPLKIAWKDGKRIKHGERVQMDFLQDGRASLRIPVLPDEGIYTAFAFNKIGNAIC 1329
QY 31 ----- 30
Db 1330 SGKLYVEPAAPLIGAPTYIPTLEPVSRIRSLSPRSVSRSPIRMSPARMSPARMSPA 1389
QY 31 ----- 30
Db 1390 RMSPCRRLEETDESOLERLYKPVFLKPVSPKCLEGGOTARFDLKVGRPMPETTFWHDGQ 1449
QY 31 ----- 30
Db 1450 QIVNDYTHKVVIKEDGTOSLIIVPATPSDSGEMTVAQNAGRSSISVILTVEAVEHQVK 1509

```
QY 31 -----LTNLNI----- 36
Db 1510 PMFVEKLNKVNIEKSGRLEMKVRATGNPNPDI VMLKNSDIIVPHKYPKIRIEGTKGAAAL 1569
QY 37 ----- 36
Db 1570 KIDSTVSODSAWYTATAINKAGRDTRCKVNVEFEPAEPEPERKLIIPRGTYRAKEIAAP 1629
QY 37 ----- 36
Db 1630 ELEPLHLRYGOBOWEBGLDYDKEKQKPFKKKLTSLRLKRPQPAHFCECLTPISDPTWV 1689
QY 37 ----- 36
Db 1690 VEWLHDGKPLEAANRLRMINEFCYCSLDYGVAYSROSGIITCRATNKYGTDHTSATLIVK 1749
QY 37 -----GA----- 38
Db 1750 DEKSLVEESOLPEGRKGLQRIEELERMAHEGALTGVTTDOKEKQKPDIVLYPEPVRVLEG 1809
QY 39 ----- 38
Db 1810 ETARFRVRTGYPOPKNVYLNQGLIRKSRFRVRYDGIHYLDIVDCKSYDGTGEVKVTAE 1869
QY 39 ----- 38
Db 1870 NPEGVIEHKVLEIOQREDFRSLRAPRPEPFHVHEPGKLOFEVQKVDRPVDTTETKE 1929
QY 39 -----ELLR--- 42
Db 1930 VVKLKRAERITHEKVPSESELRSKFRRTTEGYEAITAVELKSRKDSYEELLRKTK 1989
QY 43 ----- 42
Db 1990 DELLHWTKELTEEEKALAECKITITPFPDKIELSPSWEAPKIPERIQSOTVQGSDA 2049
QY 43 -----DPS----- 45
Db 2050 HFRVRVVGKPDCEWYKNGVKIERSDRIYVWYWPEDNVCELVIROVTAEDSASIMVKAIN 2109
QY 46 ----- 45
Db 2110 IAGETSSHAFLVQAKLITFTOELQDVVAKEDTWTATPECETSEPFVKVWKYKDGMEVH 2169
QY 46 ----- 45
Db 2170 EGDKYRMHSDRKVHFLSILITDTSADYSCVLVEDENVKTTAKLIVEGAVVEFVKELQD 2229
QY 46 ----- 45
Db 2230 IEVPESYSGELECIUSPENIEGKWYHNDVELKSNKGYTITSRGRQNLTVKQDVKEDQGE 2289
QY 46 ----- 45
Db 2290 YSFVIDGKKTCKLXMKPRPIAILQGLSDQKCEGDI VQLEVKVLSLESVEGYWMMKDGQEV 2349
QY 46 ----- 45
Db 2350 QPSDRVHIVIDKQSHMLLIEDMTKEDAGNYSFTIPALGLSTSGRVSVYSDVITPLKDVN 2409
QY 46 ----- 45
Db 2410 VIEGTKAVLECKVSPDVTSVKVYNLDEQIKPDDRVQAIKVGTQKORLVINRTHASDEGYP 2469
QY 46 ----- 45
Db 2470 KLIIVGRVETNCNLSVEKIKIIRGLRDLTCTETQNVVFEVLSHSGIDVLWNFKDKXIKPS 2529
QY 46 ----- 45
Db 2530 SKYKIEAHGKIYKLTVLNMMKDEGKYTFYAGENMTSGKLTVAGGAISKPLTDQTVAESQ 2589
```

```
QY 46 ----- 45
Db 2590 EAVFECEVANPDSKGEMLRDGKHLPLTNNIRSESDGHRKRLIIAATKLDDIGEYTVKVAT 2649
QY 46 ----- 45
Db 2650 SKTSAKLKVEAVKIKKTLKNLTVTETQDAVFTVELTHPNVKGQVQIKNGVWLESNEKYAI 2709
QY 46 -----LGAQFRVHL--VKMV-- 58
Db 2710 SVKGTIYSLRIRKNCIAIVDES VYGFRLGRGLGASARLHVETVKIIKKPKDVTALENATVAFE 2769
QY 59 ----- 58
Db 2770 VSVSHDTVPVKWPFHKSVEIKPSDKHRLVSRKVHKLMLONISPSDAGEYTA VVGQLECKA 2829
QY 59 ----- 58
Db 2830 KLFVETLHITKMKNIIEVPETKTASFCEVSHFNVPMMKNGVEIEMSEKPIVVOGKL 2889
QY 59 ----- 58
Db 2890 HOLIIMNTSTEDSAEYTFVCGNDQVSATLTVTPIMITSMKLDINAEKDTITFEVTVNYE 2949
QY 59 ----- 58
Db 2950 GISYKWLKNGVEIKSTDKCOMRTKKLTHSLNIRNVHFGDAADYTFVAGKATSTATLYVEA 3009
QY 59 ----- 58
Db 3010 RHIEFRKHIKDIKVLEKKRAMFECEVSEPDITVOMKDDOELQITDRIKIOKEYVHRL 3069
QY 59 ----- 58
Db 3070 IPSTRMSDAGKYTVVAGNVSTAKLFVEGRDVRIRSIKKEVQVIEKQRAVVEFEVNEDDV 3129
QY 59 ----- 58
Db 3130 DAHWYKDGIEINFQVERHKYVVERRIHRMFISETRQSDAGEYTFVAGNRSSVTLVYNA 3189
QY 59 ----- 58
Db 3190 PEPPQVLOELQVPTVQSGKPARFCAMISGRPOPKISWYKEBOLLSTGFKCKFLHDGQEYT 3249
QY 59 ----- 58
Db 3250 LLLIEAFPEDAAYTCEAKNDYGVAATTSASUSVEVPSDQEMPVYPPAIITPLQDTV 3309
QY 59 ----- 58
Db 3310 TSEGOPARFQCRVSGTDLKVSHYSKDKKIKPSRPFRTMQPEDTYOLEIAEAYPEDEGTYT 3369
QY 59 ----- 58
Db 3370 FVANNAVQVSSSTANLSLEAPESILHERIEOEIEMKEPSSSPLSABEGLHSAELQLS 3429
QY 59 ----- 58
Db 3430 KINETLELLSESPVYTKFDSEKGTGPIFIKEVSNADISMGDVATLSVTVIGIPKPKIQ 3489
QY 59 ----- 58
Db 3490 WFFNGCVLLTPSADYKFPDGDHSLIILFTKLEDEGEYTCMASNDYDKTICSAYLKINSK 3549
QY 59 ----- 58
Db 3550 GEGHKDTESAVAKSLEKLGPCPPHFLKELKPIRCAOGLPAIFEYTVVGEPAPVTWTF 3609
QY 59 ----- 58
Db 3610 KENKOLCTSVYYTIIHNPNSGSTFIVNDPOREDSGLYICKAENMLGESTCAAEALLVLED 3669
QY 59 ----- 58
```

Db	3670	TDMTDFCKAKSTPEAPEDFPQPLKGPVAVELDSQEIATFVKDTILKAALITENQOL	3729	Db	4750	DGKIVVEKGRIVPGVIGLMRALTINDADTDAGTYTVTVENANNLECCSCVKVVEVIRD	4809
Qy	59	-----	58	Qy	70	-----	69
Db	3730	SYEHIAKANELSSQLPLGAQELQSILODKLTPESTRFLCINGSIHFOPLKEPSNLQL	3789	Db	4810	WLKPIRDOHVVKPCTAIPACDIANKTPNIKWFKGDEIPAPBNKTEILRDGNHLYLKI	4869
Qy	59	-----ILTEPEG-	65	Qy	70	-----	69
Db	3790	QIVQSOKTFSKEGILMPEEPETOAVLSDTEKIFPMSAMSIQINSLTVEPLKTLAEPEGN	3849	Db	4870	KNAMPEDIAEYAVEIEGKRYPAKLTGEREVELLKPIEDVTIYEKESASFDAEISEADIP	4929
Qy	66	-----	65	Qy	70	-----	69
Db	3850	YPOSSIEPPMHSYLTVSAEVLSEKENTVSDTNRQRTVLOKQEAQSAALISLQSLAEGHV	3909	Db	4930	GQWKLKGLLRPSPTCEIKAEGGRFLTLHKVKLQDAGEVLYQALNAITAILTVKEIEL	4989
Qy	66	-----	65	Qy	70	-----	69
Db	3910	ESLOSPVMISQVNYEPLVPSEHSCTSGKILIESANPLENAGODSAVRIEEGSKLRFPL	3969	Db	4990	DFAVPLKDVTVPPERQARPECVLTRBANVWSKPPDIKSSDKFDIIADGKXHLVINDS	5049
Qy	66	-----	65	Qy	70	-----	69
Db	3970	ALBEKQVLLKEHSDNVMPDQIIESKREPVAIKVQEVQGRDLSKESLLSGIPBEQR	4029	Db	5050	QPDDEGVYTAIEVEGKTSARLFTVGIRLKFMSPLEDOTVKEGETATFVCELSHEKMHVVM	5109
Qy	66	-----	65	Qy	70	-----	69
Db	4030	LNLKIOTCRALQAAVASEQGLFSEWLRIEKEVEAVNITQEPRHIMCMYLVTSKSVT	4089	Db	5110	FKNDAKLHTRSTVLISSSEKTHKLEMEVTLDDISQIKAQVKELSTALQKLEADPYFT	5169
Qy	66	-----	65	Qy	70	-----	69
Db	4090	EEVTIIIEDVPQMANLKMELRDALCAIIEEIDILTAEGPRIQOGAKTSLQEBMDSFG	4149	Db	5170	VKLHDKTAVEKOBIEILUCEVSKDVPVKWPKDGEIIVSPKYSIKADGLRILKIKKADLK	5229
Qy	66	-----	65	Qy	70	-----	69
Db	4150	SOKVEPITEPEVESKYLISTEVSFNVQSRVKYLDATPVTGVASAVSDEKODESLKP	4209	Db	5230	DKGEYVDCGTDKTKANVTVEARLIEVEKPLYGVEVFGETAHFEIELSEPDVHGOWKLK	5289
Qy	66	-----	65	Qy	70	-----	69
Db	4210	SEEKESSSESGTEEVATVKIOEAGGLIKEDGPMIHTPLVDTVSEGDIVHLTTSITNA	4269	Db	5290	GQPLTASPDCEIIEDGKXHLILHNCQLGTMTEVSPQAANAKSAANLKVKELPLFITPL	5349
Qy	66	-----	65	Qy	70	-----	69
Db	4270	KEVNWYFENKLVPSDEKFKLODQNTYTLVDKVNTEDHQGEYVCEALNDSGKTATSACL	4329	Db	5350	SDVKVPEKDEAKFECEVSRPEPFTRWLKGTOEITGDDRFELIKDGTKXSHVKSAAFEDE	5409
Qy	66	-----APNI-----	69	Qy	70	-----	69
Db	4330	TVVXBAAPVIKRKTEPLEVALGHLAKFTCEIOSAPNVRFQWFKAGREIYESDKGSISSK	4389	Db	5410	AKYMFADCHTSKGLIIEGIRLKLPLTKDVTAKESAVFTVELSHDNIRVKWPKNDQ	5469
Qy	70	-----	69	Qy	70	-----	69
Db	4390	YISSLEILRTQVDCGEYTCASNEYSGVSCATLTVTVPGEKKVRKLLPERKPEKEE	4449	Db	5470	RLHTRSVSMQDEGKTHSITFKOLSIDDTSQIRVEAMGMSSEAKLTVLEGDPYFTGKLQD	5529
Qy	70	-----	69	Qy	70	-----	69
Db	4450	VVLKSVLRKRPEEPEKVEPKKLEKVKPAPVPEPPPKPVEVEVPTVTKERKIPEPTK	4509	Db	5530	YTGVEKDEVILQCEISKADAPVKWPKDGEIKPSKNVITKDGKKRMLILKALKSDIGQ	5589
Qy	70	-----	69	Qy	70	-----	69
Db	4510	VPEIKPAIPLPAPKPKPEAEVKTIKPPVPEPEPTIAAPVTVPVVGKAEAKAPKEA	4569	Db	5590	YTCDCGTDKTSGLDIEDREIKLVRPLHSVEVMEETARFETEISEDDIHANWKLKGEAL	5649
Qy	70	-----	69	Qy	70	-----	69
Db	4570	AKPGPIKGVPKKTPSIEAERRKLRPGSGGKPPDEAPPTYQLKAVPLKFVKEIKOIL	4629	Db	5650	LQTPDCEIKBEGKIHSLVHLNCRLDQGTGGVDFQAANVKSSAHLRVKPRVIGLLRPLKDVT	5709
Qy	70	-----	69	Qy	70	-----	69
Db	4630	TESEFVGSSAIFECVLSPSTAITTWKDGSNIRSPKHFADGDKRKLHIIDVOLSAG	4689	Db	5710	VTAGETATFDCELSYEDIPVEWYLGKLEPDKVPRSEGVKHTLTLRDVKLESDAGEVQ	5769
Qy	70	-----	69	Qy	70	-----	69
Db	4690	EYTCVLRGNKEKTSTAKLVVEELPVRVFKTLEEEVTVVKGQPLYLSCELNKERDVWRK	4749	Db	5770	LTAKDFKTHANLKVKEPPVEFTKPLEDOTVEEGATAVLECEVSRENAKVKWFKNGTEILK	5829
Qy	70	-----	69	Qy	70	-----	69
Db	5830	SKKYEIVADGRVKLVIHDCPTPEDIKTYTCDAKDKFTSCNLNVVPPHVEFLRPLTDQVR	5889	Db	5830	SKKYEIVADGRVKLVIHDCPTPEDIKTYTCDAKDKFTSCNLNVVPPHVEFLRPLTDQVR	5889

Qy	70	-----	69	Qy	77	-----	76
Ds	5890	EKEMARFECELSRENAKVKFKDGAIEIKKKYVDYIISKGAVALVINKLLDDEAEYSC	5949	Ds	6970	LKLENRVKTIISGEIDVNVNVIARPSAPKELFGDITKDSVHLTWBPPDDGGSPLTGVVVEK	7029
Qy	70	-----	69	Qy	77	-----LLSVC-----	81
Ds	5950	VRTARTSGMLTVLEBEAVFTKLANIEVSETDTIKLVCEVSKPGAIEVIWYKGBEIIETG	6009	Ds	7030	REVSRRKTKVMDFVTDLEFTVPDLVQGEYLFKVCARNKCGPGEPAYVDEPNMSTPAT	7089
Qy	70	-----	69	Qy	82	-----	81
Ds	6010	RYEIIITEGRKRILVIONAHLEDAAGNYNCRLPSSRTDGKVKVHLEAAEFISKQNLLEILEG	6069	Ds	7090	VPDPENVKWRDRTANSIFLTWDPKNDGSGRIKGYIVERCPRGSDKWVACGEPVAETKM	7149
Qy	70	-----	69	Qy	82	-----	81
Ds	6070	EKAIEFVCSISKESFPQMKRDDKILESGDKYDVIADGKKRVLVVKDATLQDMGTVMVVG	6129	Ds	7150	EVTGLEEGKWYAYRVKTLNROGASKSPRTEBIOAVDTQEAPEIFLDVKLLAGLTVKAGT	7209
Qy	70	-----	69	Qy	82	-----	81
Ds	6130	AARAAHLTVIEKLRIVVPLKDRVKEQOEVVFNCEVNTGAKAKWFRNEEAIFDSSKYI	6189	Ds	7210	KIELPATVGTGPEPKITWTAKADMILKQDKRITIENVPKSTVTIVDSKRSOTGTVIIIEAV	7269
Qy	70	-----	69	Qy	82	-----	81
Ds	6190	ILQKOLVYTLRIRDAHLDDQANYVSLTNHNGENVKSAANLIVEBEDURIVEPLKDIETM	6249	Ds	7270	NVCGRATAVENVVLDKPGPPAAFDITDVTNESCLLTWNPPRDDGSGKITNVVVERRATD	7329
Qy	70	-----	69	Qy	82	-----	81
Ds	6250	EKKSVTFWCKVNRNLVTLKWTNGEEVFPDNRVSVRVKYKHLMTIKDCGFPDGEYIVT	6309	Ds	7330	SEVWHKLSSTVKDTNFKATKLIIPNKEYIFRVAENMYGAGEPVQASPIAKYQFDPGPP	7389
Qy	70	-----	69	Qy	82	-----	81
Ds	6310	AGQDSVAELLIIIEAPTEFVEHEDQTVTFDDAVFSCQLSREKANVKWYNRNGREIKEG	6369	Ds	7390	TRLEPSDITKDAVTLTWCEPDDGGSPITGYWVERLDPDTPDKWVRNCNMKPVKDTTYRVKG	7449
Qy	70	-----	69	Qy	82	-----	81
Ds	6370	KYFKFGSGIHLIIKDCRLDDEYACGVEDRKSARLFBVEIIPVEIIRPPQDILEAPG	6429	Ds	7450	LTNKKYRFRVLAENLAGPGKPSKSTEPILIKDIPDPWPFGKPTVKDVGKTSVRLNWK	7509
Qy	70	-----	69	Qy	82	-----	81
Ds	6430	ADVFLAELNKQKVEQWLRNMNVVQGDQKQHMSEGKIHRLQICDIKPRDQGEYRFTAK	6489	Ds	7510	PEHDGAKIESVVIEMLKGTGDEWVRVAEGVPTTOHLLPGLMEGOEYSFRVAVNKAES	7569
Qy	70	-----	69	Qy	82	-----	81
Ds	6490	DKEARAKLEAAAPKIKTADQDLVVDVGKPLTWVPYDAYPKAAEAWFKENEPLSTKTID	6549	Ds	7570	EPSEPSDVLCREKLYPPSPPRWLEVINITKNTADLKWTVPEKOGGSPITNIVEKRDVR	7629
Qy	70	-----	69	Qy	82	-----	81
Ds	6550	TTAEQTSFRILEAKGDKGRYKIVLQNHKGKAEGFINKLVDPGVRNLEVTFDGEV	6609	Ds	7630	RKGQTVDTTVKDTKCTVTPLTEGSLYVFRVAANAIGQSDYTEIEDSVLAKDTFTTGP	7689
Qy	70	-----	69	Qy	82	-----	81
Ds	6610	SLAWEPLTDGSKIIGYVVERDDIKRKTWVLATDRAESCEFTVTGLQKGGVEYLFVSA	6669	Ds	7690	PYALAVVDVTKRHVDLKWEPKNDGGRPIORVIEKLERLCTRWVKACKTAGPCNPRVT	7749
Qy	70	-----	69	Qy	82	-----	81
Ds	6670	RNRVGTGPEVTDNPVEARSKYDVPCLNVTITDVRNFGVSLTWEPEYDGAETNVV	6729	Ds	7750	DVIEGTEVOFQVRAENEAGVGHPSPTSEILSIEDPTSPPLDLHLVTDAGRKHIAIAWK	7809
Qy	70	-----	69	Qy	82	-----	81
Ds	6730	IELRDKTSIRMDTAMTVRAEDLSATVTDVVEGQEYSFRVRAQNRIGVKGPSAATPFVKA	6789	Ds	7810	PPEKNGGSPIIIGYHVEMCPGTEKMRVNSRPIKDLKPKBEGVVPDKKEYLVRVAVNMI	7869
Qy	70	-----TANLTSS-----	76	Qy	82	-----	81
Ds	6790	DPIERPPPNLTSSDQTSVOLKWEPLKDGSPILGYIIERCEEKDNWIRCNMKLV	6849	Ds	7870	GVSEPSEISENVAKDPCKPTIDLETHDIIIVIEGEKLSIPVPPRAVPVPTVSMHKGKE	7929
Qy	77	-----	76	Qy	82	-----	81
Ds	6850	PELTYKVTGLEKNKYLVRVAENKAGVSDPSEILGPLTADDAFVEPTMDLSAFKDGLEV	6909	Ds	7930	VKASDRLTMKNDHISAHLEVPKSVRADAGIYITITLENKLGSAATINVKVIGLPGCKDI	7989
Qy	77	-----	76	Qy	82	-----	81
Ds	6910	IVPNPITILVSTGPRPTATWCFGDKVLETGDRVQMKTL SAYAELVISPSRSDKIYI	6969	Ds	7990	KASDITKSSCKLTWEPPEFGGTPILVYLERREAGRTYIPVMSGENKLSWTVKDLIPN	8049
Qy	77	-----	76	Qy	82	-----	81

Db	8050	GEYFRVAVNKVGGGEYIELKNPVIADQPKQPPDPVDEVHNPTAEAMTITWKPLDYD	8109
Qy	82	-----	81
Db	8110	GGSKIMGYIIIEKIAKEERWKRNEHLVPILTYTAKLEGEKEYQFVRAENAAGISEPS	8169
Qy	82	-----	81
Db	8170	RATPPTKAVDPIDAPKVLILTSLEVVRGDEIALDASISGSPYTIITWIKDENVIPEEK	8229
Qy	82	-----	81
Db	8230	KRAAPLVRRKGEVEEPFVLPLOTLSIDNSKKGESQLRVDRSLRPOHGLYWKVEND	8289
Qy	82	-----	81
Db	8290	HGIAKACTVSVLDTGPPINFVFEDIRKTSVLCKWEPLDDGGSEIINYTLEKKDKTKP	8349
Qy	82	-----	81
Db	8350	DSEWIVVTSLRHCKYSVTKLIEGKEYLFRVRAENRFGPGPPCVSKPLVAKDPFGPPDAP	8409
Qy	82	-----	81
Db	8410	DKPIVEDVTSNMLVKWNEPKDNGSPILGYWLEKREVNSTHWSRVNKSLLNALKANVDGL	8469
Qy	82	-----	81
Db	8470	LEGLTYVFRVCAENAAGPGKSPSPDKTAHDPISPGPPPIPRVTDTSSTIELEWEPPA	8529
Qy	82	-----	81
Db	8530	FNGGGEIVGYVDKQLVGTNKNWRCCTEKMIKRVQYTVKEIREGADYKLRVSAVNAAGEP	8589
Qy	82	-----	81
Db	8590	PGETQPVTAEPQEPFAVELDVSKGQIQIMAGKTLRIPAVVTGRPVPTKVTWKEGELD	8649
Qy	82	-----	81
Db	8650	KDRVIDNVGTSELIIKDALKRDHGRYVITATNSCGSKFAAARVEFVDPGVLDPV	8709
Qy	82	-----	81
Db	8710	VTNRKMCLNWSDEDDGGSEITGFIERDKAKMHTWRQPIETERSKCDITGLLEGOEYK	8769
Qy	82	-----	81
Db	8770	FRVIAKNKFGGPPVEIGPILAVDPLGPPTSPELTYTERQSTITILDWKEPRNGGSP	8829
Qy	82	-----	81
Db	8830	QGYIIEXRRHKDPDFERNKRLCPTTSLVENLDEHQMYEFRVKAIVEGESEPSPLNV	8889
Qy	82	-----	81
Db	8890	VIQDEVPPTIKLRLSVRGDTIKVAGEPVHIPADVTGLPMPKIENSKNETVIEKPTDAL	8949
Qy	82	-----	81
Db	8950	QITKEEVSREAKTELSIPKAVREDKGTYYTVTASNRLGVSFRNVHVEYDRSPPRNLAV	9009
Qy	82	-----	81
Db	9010	TDIKAESCYLTWDAPLDNGGSEITHYVIDKRDAKRKAEMEBVTNTAVEKRYGIWKLIPI	9069
Qy	82	-----	81
Db	9070	QOYEFVRVAVNKYGISDECKSDKVIQDPYRLPGPGPKVLARTKGSMLVSWTPPLDNG	9129
Qy	82	-----	81
Db	9130	GSPITGYWLEKREEGSPYWSRVSRAPITKVGLKGVEFNVRPLLEGVKYQFRAMAINAACI	9189
Qy	82	-----GW----- 	83
Db	9190	GPPSPSPDEVAGDPIFPGPSPCPEVKDKTKSSISLGHKPKPAKDGGSPIKGYIVEMQOE	9249
Qy	84	-----	83
Db	9250	GTTDMKRVNEPKLITTCCEVCVVPNLKELRYFRVKAIVEGESEPSDTTGEIPATDIOE	9309
Qy	84	-----	83
Db	9310	EPEVFIDIGAQCLVCKAGSOIRIPAVIKGRPTPKSSWEFDGKAKAMKDGVDHPEDAQ	9369
Qy	84	-----	83
Db	9370	LETAENSSVIIITECKRSHTGKYSITAKNAGOKTANCRVKVMDVPGPKDLKVSIDTRG	9429
Qy	84	-----	83
Db	9430	SCRLSKMPPDDGGDRIKGYVIEKRTIDGKAWTKVNPDCGSTTFVVPDLLEQQYFFVR	9489
Qy	84	-----	83
Db	9490	AENRFGIGPPVETIORTTARDPIYPPDPPIKLIKGLITKNTVHLSWPKPNDDGSPVTHY	9549
Qy	84	-----	83
Db	9550	IVECLAWDPTGKKEAWROCNRDVEELOFTVEDLVEGGEYEFBRVKAIVNAAGVSKPSATV	9609
Qy	84	-----	83
Db	9610	GPCDCQRPDMPPSIDLKEPMEVEEGTNVNI VAKIGVPPPTLTWFKAPPKPKDNKEPVLV	9669
Qy	84	-----	83
Db	9670	DTHVNKLVDVDTCLVIPOSRRSDTGLYTTITAVNNLGTASKEMRLNLVLRPGPPVPIKF	9729
Qy	84	-----	83
Db	9730	ESVSADOMTSLWFPKDDGSGKITNVVIEKREANRKTWVHVSSEPKECTYTIPTKLEGHE	9789
Qy	84	-----SOTIN----- 	88
Db	9790	YVFRIMAQNKYGIGELDSEPETARNLFSVPGAPDKPTVSSVTRNSMTVNWEEPEYDGS	9849
Qy	89	-----	88
Db	9850	PVTGYWLEMDTTSKRWKRVNRDPIKAMTLGVSYKVTGLIEGSDYQFVYAINAAGVGA	9909
Qy	89	-----	88
Db	9910	SLPSDPATARDPIAPGPPFPKVDWTKSADLENSPPLKDGSKVTGYIIVEYKEGKEE	9969
Qy	89	-----	88
Db	9970	WEGKDKEVRGTVVTVGLKEGAFYKFRVSAVNIAGIGSPGEVTDVIMKDRLVSPDLOL	10029
Qy	89	-----	88
Db	10030	DASVRDRIVVHAGGVIRIIAYVSGKPPPTVTMMNERTLPOEATIIETTAISSMWIKNCQ	10089
Qy	89	-----PEDD----- 	92
Db	10090	RSHQGVYLLAKNEAGERKKTIIIVDVLDPGPGVTGTFFLAHNLNTESKLTFSPEDDGS	10149
Qy	93	-----	92
Db	10150	PITNYVIEKRESDRRAWTPVTYVTRQNAVQGLIOCKAYFFRIAAENSIGMGPVETSE	10209
Qy	93	-----	92
Db	10210	ALVIREPITVPERPEDLEVKEVTNTVTLTNPPKYDGGSEIINYVLESRLITGTEKPHKV	10269

Qy	93	-----	92	Qy	102	-----	101
Db	10270	TNDNLLSRKYTVKGLKEGDTYEVRSVAVNI	VGCGKPSFCTKPTCKDELAPPTLHLDPD	Db	11350	TAQWNRINKDELIROCAFVRPGLIEGNEYRPR	IKAAINVGESEPRELAESVIADILHPP
Qy	93	-----	92	Qy	102	-----	101
Db	10330	KLTIRVEAFALTGRYSGPKPKVSWFKDEADV	LEDRTHTKTTPATLALIKAKRSDS	Db	11410	EVELDVTCDVITVRVGOTIRILARVKGRPE	PDITWTKEGKVLVREKRVLDLQDLPVEL
Qy	93	-----	92	Qy	102	-----	101
Db	10390	GKYCVVVNSTGSRGFCQVNVVDHPGPPVGP	VSFEVTKDYVSWKPPPLDDGSGKITN	Db	11470	QIKEAVRADHGKYIISAKNSGSHAQSAIVN	VLDRPGCONLKVTNVTKENCTISWENPL
Qy	93	-----	92	Qy	102	-----	101
Db	10450	YIIEKKEVGKVMMPVTSASAKTCKVSKL	LEGKDYIFRIHAENLYGISDPLVSDSMKAK	Db	11530	DNGSEITNFIVEYRKPNQKGSIVASDVT	KRLIKANLLANNEYFRVCAENKVGVP
Qy	93	-----	92	Qy	102	-----	101
Db	10510	DRFRVPDAPQPIVTEVTKDSALVTWKNP	HGGKPIITNYILEKRETMKRWARTKDP	Db	11590	ETKTPILAINIDRPGEPENLHADKGT	FVYLKWRPPDYDGGSPNLSYHVERRLKGSD
Qy	93	-----	92	Qy	102	-----	101
Db	10570	PYTKFRVPOLLEGQYEFVRSAENEIGID	PPSPKPVFAKDPIAKPPVPVNP	Db	11650	WERVHKSIKETHYMDRCVENQIYEF	RVQTKNEGESDMVKTBEVVVVKEDLQK
Qy	93	-----	92	Qy	102	-----	101
Db	10630	NSVDLTWOPPRHGGSKILGIYIEQKVG	DEEWRANHTPESCPETKYKVTGLR	Db	11710	LSGULTVKAGDTIRLEAGVRGKPP	EVANWTKDQATDLTRS
Qy	93	-----	97	Qy	102	-----	101
Db	10690	FRVLAVNAAGESDPAHVPEPVLVKDR	LEPPELILDANNAHQHVKVGT	Db	11770	KRSDGGKYVVTATNTAGSFVAYAT	VNVLDKPGPVRLKI
Qy	98	-----	97	Qy	102	-----	101
Db	10750	FPKVTWKEDRDAPTKARIDVTPVGS	KLEIRNAAHEDGGIYSLTVENPAG	Db	11830	EIONYILEKCEKTRMVMSTYSAT	VLTRGTTVRLIEGNEYFRVRAENKI
Qy	98	-----	97	Qy	102	-----	101
Db	10810	LDKPGPRDLEVSEIRKDCSYLTWKE	PLDDGGSVITNYVVERDVASQWSP	Db	11890	VIAKTYDKPGRPDPPEVTKVSK	KEEMTVWNPPEYDGGKSITGYFLEK
Qy	98	-----	97	Qy	102	-----	101
Db	10870	SHFAKHLNEGQYLFRAAENQYGRGP	FVETPKPIKALDPLHPPGPKDLH	Db	11950	KSAIPERRMKVONLLPDHEYQFR	VKAENEIGIGEPSLPSRVRVAKD
Qy	98	-----	97	Qy	102	-----	101
Db	10930	SLVWKNPDRDGGSPITGYLVEYQ	EEGTQDWIKFTVTNLECVVTLG	Db	12010	DTTKHSITLWGKPYDGGAPIIGY	VVEMRPKIADASPDEGKRCNA
Qy	98	-----	97	Qy	102	-----	101
Db	10990	IVGLGLPDTTPIECQEKLVPPSV	ELDVKLIEGLVVKAGTTVRFP	Db	12070	LDENQYEFVRCAQONQVIGIR	PAELKEAIKPKEILEPPEIDL
Qy	98	-----	97	Qy	102	-----	101
Db	11050	DGSEIKTDEHVTETDNFSSVLT	IKNCLRRDTGEQITVSNAAGSK	Db	12130	FAIVRGRPAPKVTKVRKVIDNV	VRKGOVLDVTMAFLVIPNSTR
Qy	98	-----	97	Qy	102	-----	101
Db	11110	PTGPINILDVTFEHTISWQPPK	DDGSPVINYIVEKODTRKDT	Db	12190	KAVFVNVRLDTRGPVSDLKVS	DVTKTSCHVSWAPPENDGGS
Qy	98	-----	97	Qy	102	-----	101
Db	11170	HLOKCEYVFRVRAENKIGVPP	LDSTPTVAKHKFSPSPGKPV	Db	12250	TVTPEVKTSFHTNLPVGN	EYFRVTAVNVEYGPVTD
Qy	98	-----	97	Qy	102	-----	101
Db	11230	PKSDGGSPIGYMERREVTKWVR	NKTPIDLKFRTVGLYEGNTY	Db	12310	ATEMTKNSATLAWLPPLR	DGAKIDGYIIISYREBEQ
Qy	98	-----	101	Qy	102	-----	101
Db	11290	KPSPSDPIKACRPIKPPGPPIN	PKDKSRETADLVMTKPSD	Db	12370	KKYKFRVAARNAGVSLPRE	AEVYEAKEQLPPKILMPEQ
				Qy	102	-----	101

Db	12430	HPCTCKWKGDEVTSSH LA VH KADSS I I I I I KDVTRKDSGYSLTAENSSGTD TQ K I V	12489
Qy	102	-----	101
Db	12490	VMDAPGPPPPF D I S I D A D A C S L S M H I P L E D G S N I T N V I E K C D V S R G D W T A L A S V	12549
Qy	102	-----	101
Db	12550	T K T S C R V K L I P G O E Y I F R V R A E N R F G I S E P L T S P K M V A Q P F G V P S E P K N A R V T K N K D	12609
Qy	102	-----	101
Db	12610	C I F V A W D R P D S G S P I I G Y L I E R K E R N S L L W K A N D T L V R S T E Y P C A G L V E G L E Y S F R I	12669
Qy	102	-----	101
Db	12670	Y A L N K A G S S P S K P T E Y T A R M P V D P P K P E V I D V T K T S V S L I W A R P K H D G S K I I G Y F V	12729
Qy	102	-----	101
Db	12730	E A C K L P G D K W R C N T A P H Q I P O E E Y T A T G L E E K A Q Y O F R A I A R T A V N I S P S E P S D P V T I	12789
Qy	102	-----	101
Db	12790	L A E N V P P R I D L S V A N K S L L T V K A G T N V C L D A T V E G K P M P T V S W K K D G T L L K P A E G I K W A M	12849
Qy	102	-----	101
Db	12850	Q R N L C T L E L F S V N R K S D V T I T A E N S S G S K S A T I K L V L D K P G P P A S V K I N K M Y S D R A M	12909
Qy	102	-----	101
Db	12910	L S W E P P L E D G S E I T N Y I V D K R E T S R P N W A Q V S A T V P I T S C S V E K L I E G H E Y Q F R I C A E N	12969
Qy	102	-----	101
Db	12970	K Y G V D P V F T E P A I K A P Y D P P G R C D P P V I S N I T K D H M T V S W K P P A D D G S P I T G Y L L E K	13029
Qy	102	-----	101
Db	13030	R E T Q A V N W T K V N R K P I I E R T L K A T G L O E G T E Y E P R V T A I N K A G P K P S D A S K A A Y A R D P Q	13089
Qy	102	-----	101
Db	13090	Y P P A P A P K Y Y D T T R S S V S L S W G K P A Y D G S P I I G Y L V E V K R A D S N W V R C N L P Q N L O K	13149
Qy	102	-----	101
Db	13150	T R F E V T G L M E D T Q Y Q F R V A V N K I G Y S D P S D V P D K H Y P K D I L I P P E G E H D A D L R K T L I L R	13209
Qy	102	-----	101
Db	13210	A G V T M R L Y P V K G R P P K I T W S K P N V N L R D R I G L I K S T D F T P L R C E N V N K Y D A G K Y I L	13269
Qy	102	-----	101
Db	13270	T L E N S C G K K E Y T I V V K V L D T G P P I N V T V K E I S K D S A Y V T W E P P I I D G S P I I N V V V Q K R	13329
Qy	102	-----	101
Db	13330	D A E R K S N S T V T T E C S K T S F R V P N L E E G K S Y F F R V F A E N E Y I G D P G E T R D A V K A S Q T P G P	13389
Qy	102	-----	101
Db	13390	V V D L K V R S V S K S C S I G W K P H S D G S R I I G Y V V D F L T E E N K W Q R V M K S L S L Q S A K D L T	13449
Qy	102	-----	101
Db	13450	E G K E Y T F R V S A E N E N G E T P S E I T V A R D D V V A P D L K G L P D L C Y L A K E N S F R L K I P I	13509
Qy	102	-----	101
Db	13510	K G K P A P S V S W K G E D P L A T D T R V S V E S S A V N T T L I V D C Q K S D A G K Y T I T L K N V A C T K E G	13569
Qy	102	-----	101
Db	13570	T I S I K V V G K P I G T P I K F D E V T A E A M T L K W A P P K D D G S E I T N Y I L E K R D S V N N K W V T C	13629
Qy	102	-----	101
Db	13630	A S A V Q K T T F R V T R L H E G M E Y T F R V S A E N K Y G V G E G L K S E P I V A R H P P D V P D A P P P N I V D	13689
Qy	102	-----	101
Db	13690	V R H D S V S L T W T D P K K T G G S P I T G Y H L E F K E R N S L L W K R A N K T P I R M R D F K V T G L T E G L E Y	13749
Qy	102	-----	101
Db	13750	E P R V M A I N A G V K P S L P S E P V A L D P I P P G K P E V I N I T R N S V T L I W T E P K Y D G H K L T	13809
Qy	102	-----	101
Db	13810	G Y I V E K R D L P S K S W M K A N H V N P E C A F T V T D L V E G G K Y E F R I R A K N T A C A I S A P S E S T E T	13869
Qy	102	-----	101
Db	13870	I I C K D E Y A P T I V D P T I K D G L T I K A G D T I V L N A I S I L G K L P K S W S K A G K D I R P S D I T	13929
Qy	102	-----	101
Db	13930	Q I T S T P S S M L T I K Y A T R K D A G E Y T I T A T N P F G T K V E H V K V T V L D V G P P G P V E I S N V S A	13989
Qy	102	-----	101
Db	13990	E K A T L T W T P L S D G S P I K S Y I L E K R E T S R L L W T V V S E D I Q S C R H V A T K L I O G N E Y I F R V	14049
Qy	102	-----	101
Db	14050	S A V N H Y G K G P V Q S E P V K M V D R F G P P P K E P S V N T K N T A T V S W K R P V D D G S E I T G Y	14109
Qy	102	-----	101
Db	14110	H V E R R E K S L R W R A I K T P V S D L R C K V T G L O E G S T Y E F R V S A E N R A G I P P S E A S D S V L M	14169
Qy	102	-----	101
Db	14170	K D A A Y P P P S N P H V T D T T K S A S L A W K P H Y D G G L E I T G Y V V E H Q V G D E A W I K D T T G T	14229
Qy	102	-LYITRF-----	107
Db	14230	A L R I T Q F V V P D L Q T K E K Y N F R I S A I N D A G V G P A V I P D V E I V E R E M A P D F E L D A E L R R T L	14289
Qy	108	-----	107
Db	14290	V V R A G L S I R I F V P I K G R P A P E V T W K D N I N L K N R A N I E S T F T L I I P E C N R Y D T G K F V	14349
Qy	108	-----	107
Db	14350	M T I E N P A G K S G F V N V R V L D T P G V L N L R P T D I T K D S V T L H W D L P L I D G S R I T N V I V E K	14409
Qy	108	-----	107
Db	14410	R E A T R K S Y S T A T T K C H K C T Y K V T G L S E G C E Y F F R V M A E N E Y G I G E P T E T T E P V K A S E A P S	14469
Qy	108	-----	107
Db	14470	P P D S L N I M D I T K T S V S L A W K P K H D G S K I T G V I E A Q R K G S D W T H I T T V K G L E C V R N	14529
Qy	108	-----	107
Db	14530	L T E G E E Y T F Q V M A V N S A G R S A P R S P V I V K S Q T M L P E L D L R G I Y Q K L V I A K A G D N I K V E	14589
Qy	108	-----	107
Db	14590	I P V L G R P K P T V T W K G D I L K Q T O R V N F E T T A T T I L I N E C V R S D S G P Y P L T A R N I V G E	14649

Qy	108	-----	107
Db	14650	VGDVITIQVHDIPGRPTGIKFEVSSDFVTSWDPENDGGVPISNVYVEMRQDSTTW	14709
Qy	108	-----	107
Db	14710	VELATTVIRTYKATRLTTGLEQFRVKAQNRVGVGGITSAMIYANYPKVPGPPTQP	14769
Qy	108	-----	107
Db	14770	VTAVTKDSMTISWHPEPLSDGGSPILGYHVERKERNILMQTVSKALVPGNIFKSSGLTDG	14829
Qy	108	-----	107
Db	14830	IAYEFRVAENMAGSKSPSEPLALDPI DPPGKVPVLNITRHTVTLKMAKPEVTGGF	14889
Qy	108	-----	107
Db	14890	KITSYIVEKRDLPNRGLKANFNLNENFTVSGLTEDAAYEFRVIAKNAAGAI SPPSEP	14949
Qy	108	-----	107
Db	14950	SDAITCRDDVEAPKIKVDVKFQDVTILKAGEAFRL EADVGRPPPTMEWSKDGKEGTA	15009
Qy	108	-----	107
Db	15010	KLEIKIADFSTNLVNKSTRRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPEGLAYTE	15069
Qy	108	-----	107
Db	15070	VTSEKCVLSWPPPLDDGAKIDHYIVQKRETSRLAWTNVASEVQVTKLVTKLLKGNEYI	15129
Qy	108	-----	107
Db	15130	FRMAVNKYGVGEPELESEPVLANPYGPPDPKPNBEVTTITKDSMVVCWGHDPDSGGSEI	15189
Qy	108	-----	107
Db	15190	INVIVERDKAQORWIKCNKKTLDLRYKVSGLTEGHEVEFRIMAENAAGISAPSTSPF	15249
Qy	108	-----	107
Db	15250	YKACDVFKPGPPGNPRVLDTSRSSISIAWKNPIYDGGSEITGYMVEIALPEDEWQIVT	15309
Qy	108	-----	107
Db	15310	PPAGLKATSYTITGLTENQEKIRIYAMNSEGLGEPALVPGTPKAEDRMLPPEIELDADL	15369
Qy	108	-----	107
Db	15370	RKVVTIRACCTLRLFPVPIKGRDPPEVKWARDHGESLDKASIESASSYTLII VGNVNRFD	15429
Qy	108	-----	107
Db	15430	GKYILTVENSGSKSAFVNVRVLDTPGPPQDLKVEVTKTSVTLTWDPDLLDGGSKIKNY	15489
Qy	108	-----	107
Db	15490	IVEKRESTRKAYSTVATNCHKTSWKVDQLQEGCSYFRVLAENEYIGILPAETAESVKAS	15549
Qy	108	-----	107
Db	15550	ERPLPGKITLMDVTRNSVLSWEKPEHDGSRILGIYIVEMQTKSGDKWATCATVKVTEA	15609
Qy	108	-----	107
Db	15610	TITGLIQEEYSFRVSAQNEKGISDPRQLSVPEVIAKDLVIPPFAKLLFNTFTVLAGEDLK	15669
Qy	108	-----	107
Db	15670	VDVPFGRPTPAVTHKONVPLKQTTTRVNAESTENNSLLTIKDACREDVGHYVVKLTNSA	15729
Qy	108	-----	107
Db	15730	GEAETLNVILDKPGPPTGPVKMDEVTA DSI TLSWGPVKYDGGSSINNYIVEKRDTSIT	15789
Qy	108	-----	107
Db	15790	TWQIVSATVARTTIKACRLKTGCEYQFRIAAENRYGKSTYLNSEPTVAQYPPKVPGPPTG	15849
Qy	108	-----	107
Db	15850	PVVTLLSRDSMEVQWNEPI SDGSRVIGYHLERKERNILMWKLNKTPITPOTKFKTTGLE	15909
Qy	108	-----	107
Db	15910	EGVEYFRVSAENIVGIGKPSKVS ECVARDPCDPPGRPEAII VTRNSVTLQWKKPTVDG	15969
Qy	108	-----	107
Db	15970	GSKITGYIVEKKELPEGRMMKASFTNII DTHFEVTGLVEDHRYEPRVIA RNAAGVFSEPS	16029
Qy	108	-----	107
Db	16030	ESTGNAITARDEVDPRI SMDPKYKDTIVVHAGESFKVDADIYKDPITQIWKDQELSN	16089
Qy	108	-----	107
Db	16090	TARLEIKSTDFATSLSVKDAVRVDSGNVILKAKNAGERSVTNVNKLDRPGPPEGVVI	16149
Qy	108	-----	107
Db	16150	SGVTAECTLAWKPPLODGGSDIINYIVERRET SRLVWTVVDANVQTLSCVKTKLEGNE	16209
Qy	108	-----	107
Db	16210	YTFRIMAVNKYGVGEPELESEPVAKNPFVVPDAPKAPVTTVTKDSMI VVHERPASDGS	16269
Qy	108	-----	107
Db	16270	EILGVLEKRDKEGIRWTRCHKRLIGELRLRVGTGLIENHDYFRVSAENAAGLSEPPS	16329
Qy	108	-----	107
Db	16330	AYQACDPIYKPGPPNNPKVIDITRSSVFLSWSKPIYDGGCEIOGYIVEKCDNVGEWTM	16389
Qy	108	-----DL 109	
Db	16390	CTPPTGINKTNIEVEKLEKHEYNFRICAINKAGVGEHADVPGPPIIVEEKL EAPDIDL	16449
Qy	110	EL	111
Db	16450	ELRKIIINIRAGSLRFLVPIKGRPTPEVKMGKVDGEIRDAAIIDVTSSTSLVLDNVNRY	16509
Qy	112	-----	111
Db	16510	DSGKYTLTLENSGTSKSAFVTVRVLDTPSPPNLKVTEITKDSVSIWPEPPLDGGSKIK	16569
Qy	112	-----	111
Db	16570	NYIVEKREATRKSAAVVTNCHKNWKIDQLQEGCSYFRVTAENEYIGILPAQTADPIK	16629
Qy	112	-----	111
Db	16630	VAEVPQPPGKITVDVDTNRNSVLSWTKPEHDGSGKIIQYIVEMQAKHSEKSECARVKS	16689
Qy	112	-----	111
Db	16690	QAVITNLTOGBEYLFVRVAVNEKGRSDPRSLAVPIAKDLVIEPDVKPAPFSSYVQGD	16749
Qy	112	-----	111
Db	16750	LKIEVPISGRPKPTITWTKDGLPLUKQTTTRINVTSLSLTTLSIKETHKDDGGQGYITVAN	16809
Qy	112	-----	111

Db	16810	VWGKTASIEIVTLDPKPPKPVKFDVSAESITLSNBPPLYTGGCQIINYIVQRD	16869
Qy	112	-----	111
Db	16870	TTVWDVVSATVARTTLKVTCLKTGTEYQFRIFAENRYGQSPALSDPIVAQYPYKEGPP	16929
Qy	112	-----	111
Db	16930	GTPFATAISKDSMVIOMHEPVNNGSPVIGYHLERKERNISILMTKVNKTIHDTQPKAQN	16989
Qy	112	-----	111
Db	16990	LEEGIEYFRVYAENIVGVGKASKNSECYVARDCDPPGTPPEPIMVKRNEITLQWTKPVY	17049
Qy	112	-----	111
Db	17050	DGGSMTICYIVEKRDLPDGRMMKASFTNVLETQFTVSLGTEDQRYEFRVIAKNAAGAIK	17109
Qy	112	-----	111
Db	17110	PSDSTGPITAKDEVELPRISMDPKPRDTIVVNAGETFRLEADVHGKPLPTIEWLRGDKEI	17169
Qy	112	-----	116
Db	17170	EESARCEIKNTDFKALLIVKDAIRIDGGQYILRASNVAGSKSFPPNVKVLDRPPEGPV	17229
Qy	117	QVRGVT-----QLGG-----	126
Db	17230	QVTGVTSEKSLTWSPPLODGGSDISHYVVEKRETSRLAWTVVASEVVTNSLKVTKLEGG	17289
Qy	127	-----	126
Db	17290	NEYVFRIMAVNKYGVGPESAPVLMKNPFVLPGPSKLEVTNIAKDSMTVCMNRPSDG	17349
Qy	127	-----	126
Db	17350	GSEIIGYIVEKRDGSRIRWIKCNKRITDLRLRVTLGTEDHEYFRVSAENAAVGEPSP	17409
Qy	127	-----ACSP-----	130
Db	17410	ATVYKACDPVKPGPPTNAHI VDTTKNSITLANGKPIYDGGSEILGVVVEICKADEEW	17469
Qy	131	-----	130
Db	17470	QIVTPOTGLRVTRFEISKLEHQBKYKIRVCALNKVGLGEATSVPGTVKPEDKLEAPELDL	17529
Qy	131	-----TWS-----	133
Db	17530	DSELRKGIVVRAGGSAR:HIPFKGRPMPEITWSREGEFTDKVQIEKGVNYTQLSIDNCD	17589
Qy	134	-----	133
Db	17590	RNDAGKYILKLENSGSKSAFVTVKVLDTPGPPQNLAKEVKDSAFVLEWPEIIDGAK	17649
Qy	134	-----	133
Db	17650	VKNYVIDKRESTRKAYANVSSKSKTSFKVENLTEGAIYYFRVMAENEFVGVVETVDA	17709
Qy	134	-----	133
Db	17710	VKAAEPPSPPGKVTLTDSQTSASLWKEKPEHDGGSRLVGVVEMQPKGTEKWSIVAESK	17769
Qy	134	-----	133
Db	17770	VCNAVVTGLSSQEQYQFRVAKYNEKSGDPRLVGVPIAKOLTIQPSLKLPTNTYSIQAG	17829
Qy	134	-----	133
Db	17830	EDLKEIPVIGRPRNISWVKDGEPLKQTRVNVBEATATSVLHIKEGNKDDFGKYTVTA	17889
Qy	134	-----	133
Db	17890	TNSAGTATENLSVIVLEKPGPPVGPVRFDEVSADFVVISWEPPAYTGGCOISNYIVEKRD	17949
Qy	134	-----	133
Db	17950	TTTTTWHMVSATVARTTIKITKLKTGTEYQFRIFAENRYGKSAPLDSKAVIVQYPPKEPG	18009
Qy	134	-----	133
Db	18010	PGOTPFVTSIKDQMLVQWHEPVNDGGTKIIGYHLEQEKNSILWKLNKTPIODTKFKT	18069
Qy	134	-----	133
Db	18070	TGLEGLEYEYFKVSAENIVGIGKPSKVSECFVARDCDPPRPEAIVITRNNVTLKWKPK	18129
Qy	134	-----	133
Db	18130	AYDGSKITGYIVEKXDLDPGRMMKASFTNVLETETVSLVEDQRYEFRVIAARNAAGNF	18189
Qy	134	-----	133
Db	18190	SEPSDSSGAIARDEIDAPNASLDPKYKDVIVVHAGETFVLEADIRGKPIPDVVMSKDGGK	18249
Qy	134	-----	133
Db	18250	ELEETAARMEIKSTIQKTTLVKDCIRTDGGQVILKLSNVGGTGSIPITVKVLDRLPGSPE	18309
Qy	134	-----	133
Db	18310	GPLKVTGVTAEKCYLAWNPPLODGGANISHYIIEKRETSRLSWTQVSTEVALNKYKTKL	18369
Qy	134	-----	133
Db	18370	LPONEYIFRMAVNVKYGIGEPLESGPVTACNPYKPPGPPSTPEVSIAITKDSMVVTVWARPV	18429
Qy	134	-----	133
Db	18430	DDGTEIEGYILEKRDKEGVRWTKCNKKTLDLRLRVTLGTEGHSYEFRAAENAAAGVGE	18489
Qy	134	-----	133
Db	18490	PSEPSVFRACDALYPPGPSNPKVTDTSRSSVSLAWSKPIYDGGAPVKGYVVEVEKAAA	18549
Qy	134	-----	133
Db	18550	DEWTTCTPTGLOGKQFTVTKLXENTENYFRICAINSEGVEGEPATLPGSVVAQERIEPPE	18609
Qy	134	-----	133
Db	18610	IELDADLRKVVLRLASATLRLFVTIKGRPEPEVKWEKAEGLTDRAOIEVTSSTMLVID	18669
Qy	134	-----	133
Db	18670	NVTRFDSGRYNLTLNENSGSKTAFVNVRLDPSAPVNLTIREVKKDSVTLSEBPLIDG	18729
Qy	134	-----	133
Db	18730	GAKITNIVIEKRETRKAYATITNNCTKTTFRIENTLOEGCSYFRVLASNEYGIGLPAET	18789
Qy	134	-----	133
Db	18790	TEPVKVSEPLPPGRVTVLDVTRNTATIKWEKPEDSGSKITGYVWEMQTKSEKWSCTCT	18849
Qy	134	-----	133
Db	18850	QVKTLEATISGLTAGEEYFRVAANVEKGRSDPRQLGVPVIARDIEIKPSVELPFHTFNV	18909
Qy	134	-----	133
Db	18910	KAREQLKIDVPFKGRPOATVNRKDGOTLKETTRVNVSSSKTVTSLSIKEASKEDVGTYE	18969
Qy	134	-----	133
Db	18970	LCVNSAGSITVPITIIIVLDRPGPPGPIRIDEVSCDSITISWNPPEYDGGCOISNYIVEK	19029

QY	134	-----	133	QY	134	-----	133
Db	19030	KETTSTTHIVSOAVARTSIKIVRLTTGSEYQFRVCAENRYGKSSYSESSAVVAEYFPSP	19089	Db	20110	LLKRETSTAVMTSATVARTMMKVMKLTGTBEOYQFRIKAENRFGISDHIDSACVTVKLP	20169
QY	134	-----	133	QY	134	-----	133
Db	19090	PGPPGTPKVHATKSTMLVTWQVPVNDGGSRVIGYHLEYKERSIILWSKANKILIADTQV	19149	Db	20170	YTPPGPSTPWNTVNTRESITVWHEPVSNGGSVAVGVHLEMKDRNSILWOKANKLVIPT	20229
QY	134	-----	133	QY	134	-----	133
Db	19150	KVSGLDEGLMYEYRVVAENIAGIGKSKSCEPVPARDCPPGQPEVTNITRKSLSLWS	19209	Db	20230	THFKVTISAGLIYEYRVVAENAGVGKPSHPSEPVLAIDACEPPRNVRITDISKNSVL	20289
QY	134	-----	133	QY	134	-----	133
Db	19210	KPHYDGAKITGVIVERRELPGDMLKCHYTNIQETYFEVTELTEQRYEYFVARNAD	19269	Db	20290	SMQOPAFDGGSKITGYIVERRDLPGRWTKASFTNVTTQFTISGLTONSOYEFVARN	20349
QY	134	-----	133	QY	134	-----	133
Db	19270	SVSEPESTGPIIVKDDVEPRVMDVKFRDVIIVKAGEVLKINADIAGRPLPVISWAKD	19329	Db	20350	AVGSISNPSEVVGPITCIDSYGGPVIDPLEYTEVVKYRAGTSVKLRAGISCKPAPTIEW	20409
QY	134	-----	133	QY	134	-----	133
Db	19330	GIEIERARTEIISTDNHTLLTVKDCIRDTGOYVLTILKNVACTRSVAVNCKVDPKGGPP	19389	Db	20410	YKDDKELOTNALVCVENTTDLASILI KQADRLNSCYELKLRNAMASASATIRVOIDKP	20469
QY	134	-----	133	QY	134	-----	CLIT 137 :
Db	19390	AGPLEINGLTAEKCSLSWGRPOEDGDIDYVHRKXRETSHLAWTICEGELQWTSCKVTK	19449	Db	20470	GPPGPIEFKTVTAEKITLLWRPPADGGAKITHYIVREKRETSRVVMSVMSVHLEBECIIT	20529
QY	134	-----	133	QY	138	-----	137
Db	19450	LLKGNEYIFRVTGNVKGVGEPLESVAIKALDPFTVPSPPTSLEITSVTKESMTLWSRP	19509	Db	20530	TTKIIKIGNEYIFRVAENVKYGIGEPLESVVAKNAFVTPGPPGPIEVTKITKNSMTVVM	20589
QY	134	-----	133	QY	138	-----	137
Db	19510	ESDGGSEISGYIIEREKNSLRVVRVKNPVYDLRVKSTGLREGCEYEVYVAENAAUGS	19569	Db	20590	SRPIADGSDISGYFLEKRDKSLGWFVKVLETIRDTQKVTGLTENSQYQYRVCVNAA	20649
QY	134	-----	133	QY	138	-----	137
Db	19570	LPSETSPLRAEDPVLPPSPSPKIVDSGKTTITIAWKVPLFDGGAPITGYTVEYKXSD	19629	Db	20650	GQPFSEPFYKAAADPIDPGPPAKIRIADSTKSSITLGWSKPYVYDGGSAVTGYVWEIR	20709
QY	134	-----	133	QY	138	-----	137
Db	19630	DTDWKTSIQSLRGTEYTIISGLTTGAEYVFRVKSVNKVGASDPSDSDPOIAKEREPEPLF	19689	Db	20710	QGESEEWTTVSTKGEVRTTEYVVSNLKPGVNYFRVSAVNCAGOGEPITEMPEVQAKOIL	20769
QY	134	-----	133	QY	138	-----	137
Db	19690	DIDSEMRKTLIVKAGASFTMTVPFRGRVPVNLWSKPDTLRTRAYVDTTDSRTSLTIEN	19749	Db	20770	EAPRIDLVALRTSVIAKAGEDVOVLI PFKGRPPPTVTRKDEKNLGDARSIENTDSS	20829
QY	134	-----	133	QY	138	-----	137
Db	19750	ANRDSGKYTLTIONVLSAASLTLVVKVLDTPGPPTNITVQDVTKESAVLSWDVPENDGG	19809	Db	20830	SILLTIPOVTRNDTGKVIILTIEINGVGEKSSTVSVKVLDTPAACQKLQVHVSRGTVLLW	20889
QY	134	-----	133	QY	138	-----	137
Db	19810	APKNVYHIEKREASKAWSVTNNCNRLSVKVTNLOEGAIYFRVSGENEFVGVIPAETK	19869	Db	20890	DPPLIDGGSPIINYVIEKRDATKRTMSVVSHKCSSTSPKILDLSEKTPFFFRVLAENEIG	20949
QY	134	-----	133	QY	138	-----	137
Db	19870	EGVKITEKPSPEKLGVTISKDSVSLTWLKPEHDGGSRIYVHVVEALEKGQKNWYKCAV	19929	Db	20950	IGEPCEETPEVKAEEVPAPIRDLMSKDKSTKTSVILSWTKPDGGSVITEYVVERKKGCE	21009
QY	134	-----	133	QY	138	-----	137
Db	19930	AKSTHHVSLRSENSEYFRVFAENQAGLSDPRELLPVLIXEOLBPPPIDMKNFPSHTV	19989	Db	21010	QTMHAGISKTCEIEVSQLEKQSVLEPRVFAKNEKGLSDPVTIGPITVKELIITPEVDLS	21069
QY	134	-----	133	QY	138	-----	137
Db	19990	YVRAGSNLKVDPISCKPLPKVTLSRDGVPLKATMFPNTEITAENLTINLKESVTADAGR	20049	Db	21070	DIPGAQVTVRIGHNVHLEPYKGPSPISMLKGLPLKESEFVRFSKTENKITLSIKNA	21129
QY	134	-----	133	QY	138	-----	137
Db	20050	YEITAANSSGTTKAFINIVULDRPGPTGPVVISDITEESVTLKWEPPKYDGGSQVTNVI	20109	Db	21130	KKEHGKGYTILDNNAVCRIAPVITVITLGPSPKGPPIRFDKADSVILSWDPEDNCG	21189
				QY	138	-----	137

Db	21190	GEITCYSIEKRETSOTNWKWVCSSVARTTFKVPNLVKDAEYQFRVRAENRGVSQPLVSS	21249
Qy	138	-----	137
Db	21250	IIVAKHQPRIQPPGKPIYVNVTDGMSLTWDAPYDGGSEVTGFHVEKKERNILQKV	21309
Qy	138	-----	137
Db	21310	NTSPISGREYRATGLVEGLDYQFRVYAENSAGLSPPDSPKFTLAVSPVDPGPDYDV	21369
Qy	138	-----	137
Db	21370	TRETITLKNPPLRDGSKIYVGSIEKROGNERWRCNFTDVSECOYTVTGLSPGDRYEF	21429
Qy	138	-----	137
Db	21430	RIIARNAVGTISPSSQSGIIMTRDENVPPIVEFGPEYFDGLIIKSGESLRKALVQGRP	21489
Qy	138	-----	137
Db	21490	VPRVTWKDGVIEIKRMNMEITNVLGSTSLFVRDATRDHRGVYVTEAKNASGSAKAEIKV	21549
Qy	138	-----	137
Db	21550	KVQDTPGVGVPIRFTNITGEMTLWWDAPLNDGCAPIITHYIIEKRETSRLAWALIEDKC	21609
Qy	138	-----	137
Db	21610	EAQSYTAIKLINGNEYQFRSAVNKFGVGRPLDSDPVVAQIQYTVDPAPGIPSPNITGN	21669
Qy	138	-----	137
Db	21670	SITLTWARPESDGGSEIOQYILERREKSTRWVKVISKRPISETRFKVKTGLEGNEYEFH	21729
Qy	138	-----	137
Db	21730	VMAENAGVGASGISRLIKREPVPNPGPPTVVKVTDTSKTTVLSWSKPVFDGMEII	21789
Qy	138	-----	137
Db	21790	GYIIEMCKTDLGDWHKVNAEACVKTRYTVTDLQAGEEYKFRVSAINGAGKDSCEVTGTI	21849
Qy	138	-----	137
Db	21850	KAVDRLTAPELDIDANFKQTHVVRAGASIRLFIAYQGRPTPTAVWSKPDNSLSURADIHT	21909
Qy	138	-----	137
Db	21910	TDSFSTLTVCNCRNDACKYTLTVENNNGSKSITFTVKVLDTPGPPGPIFTKDVTRGSAT	21969
Qy	138	-----	137
Db	21970	LMDAPLLDGGARIHHYVVEKREASRSWQVISEKTRQIFKVNDLAEGVPYFRVSAVN	22029
Qy	138	-----	137
Db	22030	EYGVGPEYEMPEPIVATEQAPPRRLDVVDTSKSAVLAWLKPDHDGGSRTGYLLENRQ	22089
Qy	138	-----	137
Db	22090	KGSDLWVEAGHTKQLTFTVERLVEKTEYEFVRKAKNDAGYSEPREAFSSVIIKEPQIEPT	22149
Qy	138	-----	137
Db	22150	ADLTGITNQLITCKAGSPFTIDVPISGRPAKVTWKLEMLRKETDRVSIITTKDRTTLT	22209
Qy	138	-----	137
Db	22210	WDSMRGDSGRYFLTLENTAGVKTFSVTVVVIGRPGPVTGPIEVSSVSAESCULSWGEPK	22269
Qy	138	-----	137
Db	22270	DGGGTEITNYIIVEKRESGTTAQLVNSSVKRTQIKVTHLTKYMEYSFRVSSSENRFQVSKP	22329
Qy	138	-----	137
Db	22330	LESAPIIAEHPFVPSPAPTRPEVYHVSAANAMIRWEEPYHDGSKIIGYVWEKKERNITIL	22389
Qy	138	-----	137
Db	22390	WVKENKVPCLCENYKVTGLVEGLEVOFRYYALNAAQVSKASEASRPIMAQNPVDAPGRPE	22449
Qy	138	-----	137
Db	22450	VTDVTRSTVSLIWSAPAYDGGKVGYYIIERKPVSEVGDRMLKCNYYTIVSONPFTVTAL	22509
Qy	138	-----	137
Db	22510	SEGDTYEFRLAKNAAGVISKSGSESTGPTVTCRDEVAPPKAELOARLHGLDLVTRAGSDLV	22569
Qy	138	-----	137
Db	22570	LDAAVGGKPEPKIINTKGDKELDLCEKVSLOYTKRATAVIKFCDRSDSGKYTLTVKNAS	22629
Qy	138	-----	137
Db	22630	GTKAVSMVMKVLDSPGCGKLTVSRVTOEKTCLAWSLPOEDGAEITHYIVERRETSRLN	22689
Qy	138	-----	137
Db	22690	WVIVEGECPTLSYVWVTRLKNNHEYIFRVAENVKYQGVPESEPIVARNSTFTIPSPGP	22749
Qy	138	-----	137
Db	22750	EEVGTGKEHIIQWTKPESDGGNEISNLYVDKREKSLRWTRVKNKYVYVDTRLKVTSLM	22809
Qy	138	-----	137
Db	22810	EGCDYQFRVTAVNAAGNSEPSPERSNFISCREPSYTPGPPSAPRVVDTTKHSISLAWTKPM	22869
Qy	138	-----	137
Db	22870	YDGGTDIVGYVLEMOEKOTDQWYRVHTNATIRNTEFTVPDLKMGOKYSPRVAANVVKGMS	22929
Qy	138	-----	137
Db	22930	EYSESTAEIEPVERIEIPDLLEADDLKKTVTIRAGASLRLMVSVSGRPPPVITWSKQID	22989
Qy	138	-----	137
Db	22990	LASRAIIDTTESVSLIIVDKVNRDAGKYTIEAENQSGKSATVLVKVYDTPGPCPSVKV	23049
Qy	138	-----	137
Db	23050	KEYSRDSVTITWEIPTIDGGAPINNNYIVEKREAAAMRAFKVTVTTKSCTLRYISGLVEGTM	23109
Qy	138	-----	137
Db	23110	HYFRVLPENIYGIGBECETSDAVLYSEVPLVPAKLEVVVDTKSTVTLAMEKPLYDGGSL	23169
Qy	138	-----	137
Db	23170	TGYVLEACKAGTERMMKVVTLPKPTVLEHTVTSLINEGEQYLFRIAQNEKGVSEPRETVTA	23229
Qy	138	-----	137
Db	23230	VTQDLRVLPITDLSTMPQKTIHVPAGRPVELVPIAGRPPPAASWFFPAGSKLRESERV	23289
Qy	138	-----	141
Db	23290	VEHTKVAKLTIRETITRDTGTYLLEKVNVTGTTSETIKVIIIDKPGPTGPIKIDEIDA	23349
Qy	142	-----	141
Db	23350	TSITISWEPPELDGGAPLSGYVVEORDAHRPGWLPVSESVTRSTFKFTRLTEGNEYVFRV	23409

```
Qy 142 ----- 141
Db 23410 AATNREGISYQSEVIECRSSIRIPGPETLQIFDVSRDGWTLTWYPEDDGGSQVTCY 23469
Qy 142 ----- 141
Db 23470 IVERKEVRADRWVRNVKVPVMTWTRYSRSTGLTEGLEYEHRVTAINARGSGKSPRSPKPIVA 23529
Qy 142 ----- 141
Db 23530 MDPIAPPGKPNPRVTDTRTSVSLAWSVPEDEGGSKVTGYLIEMQKVDQHEWTKCNTTP 23589
Qy 142 ----- 141
Db 23590 TKIREVTLTHLPOGAERYFRVLACNAGGPEPAEVPGTVKVTEMLEYDPYELDERVQEG 23649
Qy 142 ----- 141
Db 23650 FVRQGGVIRLTIPIKGKPPFICKWTKEGQDISKRAMIATSETHTELVIKEADRGDSGTVD 23709
Qy 142 ----- 141
Db 23710 LVLENKCGKANVYIKURVIGSPNSPEGPLEYDDIQVRSVRVSWRPPADGGADILGYILE 23769
Qy 142 ----- 141
Db 23770 RREVPKAAMYTIDSRVGTSLVVKGLKENVEYHFRVSAENQFGISKPLKSEEPVTPKTPL 23829
Qy 142 ----- 141
Db 23830 NPPEPPNPPELVDTKSSVSLWSRPKDDGGSVRTGYIERKETSTDKWRHNTQITT 23899
Qy 142 ----- 141
Db 23890 TWYTVGLVDAEYOFRIIAQNDVGLSETSPASEVPVCKDPDKPSQGELEILSISKDS 23949
Qy 142 ----- 141
Db 23950 VTLQEKPCDGGKEILGYWVEYRQSDSAWKSNNERIKDKQFTIGGLEATEYEFVRF 24009
Qy 142 ----- 141
Db 24010 AENETGLSRPRRTAMSIKTLATSGEAPGIRKEMKDVTTLKGEAAQLSCQIVGRPLDQKW 24069
Qy 142 ----- 141
Db 24070 YRFGKELIQSRKYKMSDGRHTLTVMTTEEQDEGVYTCIATNEVGEVETSSKLLQATP 24129
Qy 142 ----- 141
Db 24130 QPHPGYPLKEYYGAVGSTRLHLVMYIGRPVPMTWPHGQKLLQNSENITIENTHYTHL 24189
Qy 142 ----- 141
Db 24190 VMKNVQRKTHAGKYKVQLSNVFGTDAILDVEIQDPKPTGPVIEALLKNSAVISWKP 24249
Qy 142 ----- 141
Db 24250 PADDGGSWITNYVEKCEKAEBWQLVSSAISVTTCRIUNLITENAGYFFRVSAQNTFGI 24309
Qy 142 ----- 141
Db 24310 SDPLEVSSVVIKSPPEKPGAPKPTITAVTKDSCVAVKPPASDGGAKIRNYLEKREK 24369
Qy 142 ----- 141
Db 24370 KONKWSVTBETIRETFVSKNLIBGLEVEFRVKCNLGESEWSEISEPTPKSDVPIQ 24429
Qy 142 ----- 141
Db 24430 APHFKEELRNLRVQSNATLVCKVTGHPKPIVKWYROCKEIIADGLKYRIOEFKGYHQ 24489
```

```
Qy 142 ----- 141
Db 24490 LIIASVTDDATVYQVRATNOGGSVSGTASLEVEVPAKIHLPKLTLEGMAVHALRGEVVS 24549
Qy 142 ----- 141
Db 24550 IKIPFSGKPDPIITWQKQDLIDNNGHYQIVTRSFSTSLVFPNGVERKDAGFYVVCANR 24609
Qy 142 ----- 141
Db 24610 FGIDOKTVELDAVDPPDPGRGVKVSADSRDSVNLTWTEPASDGGSKIINYIVEKATTAE 24669
Qy 142 ----- 141
Db 24670 RMLRVQARETRYTVINLFGKTSYQFRVIAENKFGLSKPSBPSEPTITKEDKTRAMNDE 24729
Qy 142 ----- 141
Db 24730 EVDETREVSMTKASHSTKELYEKYMIADLGRGEFGIVHRCVETSSKTYMAKFKVKVG 24789
Qy 142 ----- 141
Db 24790 TDQVLVKEISILNIARHRNHLHLSFESMEELVMIFEPISGLDIFERINTSAPELNER 24849
Qy 142 ----- FDL ----- 144
||:
Db 24850 EIVSVHVQCEALQFLHSHNIGHDFDIRPENIIYQTRRSSTIKIIEFGOARQLKPCDNFRL 24909
Qy 145 ----- 144
Db 24910 LFTAPEYVAPEVHQHVSTATDMWSLGLTVYVLLSGINPFLAETNOOIIENIMNAEYTF 24969
Qy 145 ----- 144
Db 24970 DEEAFKEISIEAMDFVDRLLVKKRSMTASEALQHPWLKOKIERVSTKVIRTLKHRRYY 25029
Qy 145 ----- 144
Db 25030 HTLIKDLNMVVSAAIRSCGGAIRSQKGSVAKVKVASIEIGPVSGQIMHVGEGGHVK 25089
Qy 145 ----- 144
Db 25090 YVCKIENYDOSTQVTWYFVGRQLENSKEYEITYEDGVAILYVKDITKLDDGTYRCKVVD 25149
Qy 145 ----- 144
Db 25150 YGEDSSVAELFVKGVREVYDYCRRTMKKIKRRTDTRLLRPPPEFTLPLYNKTYVGEN 25209
Qy 145 ---GVTT 148
|||
Db 25210 VRFGVTI 25216

RESULT 10
US-918-171A-7
; Sequence 7, Application US/09918171A
; Patent No. US20020110894A1
; GENERAL INFORMATION:
; APPLICANT: Apce, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/04193
; CURRENT APPLICATION NUMBER: US/09/918,171A
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 09/369,364
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
```


US-09-918-171A-7

Query Match 30.7%; Score 237; DB 10; Length 997;
Best Local Similarity 18.2%; Pred. No. 0.011;
Matches 61; Conservative 24; Mismatches 62; Indels 188; Gaps 6;
QY 1 AAGGILHLEL-----PD-----VFAHQED-----LVAVG----- 15
DB 53 AGGSFLSYELMPLRALRKDVSVRRDAPAFYELQYRGLRNLNLTANQHLLAPGVSETRR 112
QY 16 -----PD-----VFAHQED-----LVAVG----- 25
DB 113 RGGIGRAHRAHTPACHLLGEVQDPELEGGAAISACDGLKGVFQLSNEDYFIEPLDSAP 172
QY 26 -----PD-----VFAHQED-----LVAVG----- 25
DB 173 ARPGHAOPVYVKRQAPERLAORGDSSAPSTCGVQVYPELESRRERWEQQRRLRR 232
QY 26 -----PD-----VFAHQED-----LVAVG----- 25
DB 233 LHQSVSKEKWCETLVADAKWYHGQPVESYVLTIMNVAGLFHDPSIGNPIHITIV 292
QY 56 KWLITPEGAPNITANLTSLLSVCGWSOTINPEDDTPGHADLVLYITRFDELDPGN 115
DB 293 RLVLDEEEDLKTTHADNTLKSCKWQKSINMKGDARPLHDDTAILLTRKDL-CAAMN 351
QY 116 R--QVRGVTOLGGACSPWMSCLITDGTGDLGVTI 148
DB 352 RPECETGLSHVAGMCQPHRSCSINEDTGLPLAFTV 386

RESULT 11

US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U51
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-6

Query Match 30.4%; Score 235; DB 9; Length 11877;
Best Local Similarity 1.1%; Pred. No. 4.1;
Matches 91; Conservative 21; Mismatches 36; Indels 8329; Gaps 25;
QY 1 AAGG-----PD-----VFAHQED-----LVAVG----- 4
DB 450 AAGGAQVTDVPADRWAGDFYDPRSAPGRSNRWGGFIEDVDRFDAFFGISPREAEM 509
QY 5 -----PD-----VFAHQED-----LVAVG----- 4
DB 510 DPOORALELGEALERAGIDPSSLTGTRGVFAGAIWDDYATLKRGGAAITPHTVTG 569
QY 5 -----PD-----VFAHQED-----LVAVG----- 4
DB 570 LHRGIIANRLSYTTLGRPSMWVDSGSSSLVAVHLACESLRRGESEALAGVSLNLP 629
QY 5 -----PD-----VFAHQED-----LVAVG----- 4
DB 630 DSIICASKFGGLSPDGRAYTTFDARANGVVRGEGGVVVKLSRAVDGDPVLAVIRGSA 689

QY 5 -----PD-----VFAHQED-----LVAVG----- 4
DB 690 VNNGAAQMTTPDAQAEAVLREAHERRAGTAPADVRYVELHGTGTPVGDPIEAAALGAA 749
QY 5 -----PD-----VFAHQED-----LVAVG----- 9
DB 750 LGTRPAGOPLLVGVSKTNIHLEGNAGTAGILKAVLAVGRALPASLNYETNPAPPE 809
QY 10 -----PD-----VFAHQED-----LVAVG----- 9
DB 810 EMLNLRVNTVEYLPWEPEHGDQRMVGVSVFGMGTTNAHVLEEBAPGCGRGASVVESTVGS 869
QY 10 -----PD-----VFAHQED-----LVAVG----- 9
DB 870 AVGGVYVPMVSAKSAALDAQIERLAFAFSRDRTDGVDAGAVDAGAVARVLGG 929
QY 10 -----PD-----VFAHQED-----LVAVG----- 17
DB 930 RAQFEHRAVVVSGGPDLLAALAAPEGLVGVASGVGRVAFVPPGCGTOMAGMGAELLS 989
QY 18 --VFQA-----PD-----VFAHQED-----LVAVG----- 21
DB 990 SAVFAAAMAECEAALSPLYVDSLEAVVROAPGAPTLEVRVDVQVTFVAVMSLARVMOHH 1049
QY 22 -----PD-----VFAHQED-----LVAVG----- 21
DB 1050 GVTPOAVVGHSGEIAAYVAGALSDDAARVVTLRSKISIAHLAKGGMLSLAUSEDV 1109
QY 22 -----PD-----VFAHQED-----LVAVG----- 21
DB 1110 LERLAGFDGLSVAANVGPTATVVSGDPVQIEELARACEADGVRRVIVPDYASHSRQVEI 1169
QY 22 -----PD-----VFAHQED-----LVAVG----- 21
DB 1170 IESELAELVAGLSPQAPRVPPFSTLEGAMITEPVLDDGGYWRNLRHRRVGFAPAVETLATD 1229
QY 22 -----PD-----VFAHQED-----LVAVG----- 21
DB 1230 EGFTHFEVSAHPVLTALPGVTGLATLRRDNGODRLVASLAERANGLAVDWSPLLP 1289
QY 22 -----PD-----VFAHQED-----LVAVG----- 28
DB 1290 SATGHHSDLPYAFQTERHNLGELIAPAGEPAVQPAVLRTENAEPALDRDEQLRVIL 1349
QY 29 -----PD-----VFAHQED-----LVAVG----- 28
DB 1350 DKVRAQTAQVLGVATGGQIEVDRTFREAGCTSLTGYDLNRINAAFGVRMAPSMIPDPT 1409
QY 29 -----PD-----VFAHQED-----LVAVG----- 28
DB 1410 PEALAEQLLLVHGEAANPAGAEPAVAAAGAVDEPVAIVGMACELPGCVASPEDLWRL 1469
QY 29 -----PD-----VFAHQED-----LVAVG----- 28
DB 1470 VAGGDAISEFPQDRGMVDEGLYHPDPEHPGTSYVRQGGFIENVAGDAAFFGISPREAL 1529
QY 29 -----PD-----VFAHQED-----LVAVG----- YVLT 32
DB 1530 AMDPOORLLLETWEAVEDAGIDPTSLRGRQVGVFTGAMTHEYGPFSLRDGGEGLDGYLLT 1589
QY 33 N-----PD-----VFAHQED-----LVAVG----- 42
DB 1590 GNTASVMSGRSVYTLGLEPALTVDTCSSSLVALHLAVQALRKGEVDNALAGGVAMPT 1649
QY 43 -----PD-----VFAHQED-----LVAVG----- 42
DB 1650 PGMFVEFSRQRLAGDRSKAFASADGTSWSEGVGVLLVERLSDARRNGHQLVAVRGS 1709
QY 43 -----PD-----VFAHQED-----LVAVG----- 42
DB 1710 ALNODGASNGLTAPNGPSQORVIRRALADARLTTSDDVVEAHGTGTRGLGDPIEAQALIA 1769

QY 43 ----- 42
Db 1770 TYGGRDEOPLRLGLSLKSNIGHTQAAAAGVSGVIKMWQAMRHGLPKTLHVDESPDQIDW 1829
QY 43 ----- DPSLG- 47
Db 1830 SAGAVELLTEAVDPKQDGLRRAAVSFGISCTNAHVLEAPVVPVVEGASVVEPSVVG 1889
QY 48 ----- AOF 50
Db 1890 SAVGGVTPVWVSAAKSAALDAQIERLAAFASRDRTDDADAGAVDAGAVAHVLADGRAQF 1949
QY 51 ----- 50
Db 1950 EHRAVALGAGADBLVQALADPDGLIRGTASGVRVAFVFPQGCTQWAGMAELLDSSAVF 2009
QY 51 ----- RVHLV- 55
Db 2010 AAMAECEAALSPPVDWSLEAVVRQAPGAPTLERVDVQVQVTPFVVMVSLARVMQHHGVTP 2069
QY 56 ----- 55
Db 2070 QAVVGHSGEIAAAYVAGALPLDDAARVVTLRSKSTIAAHLAKGGMLSLALNEDAVLERL 2129
QY 56 ----- 55
Db 2130 SDFGLSVAANGPTATVWSGDPVQIEELAQACKADGFRARIIPVDYASHSRQVEIESE 2189
QY 56 ----- 55
Db 2190 LAQVLAGLSPOAPRPVFFSTLECTWITEPVLDTGYWYRNLHRVGPAPAIETLAVDEGFT 2249
QY 56 ----- RVHLV- 55
Db 2250 HFVEVSAHPVLTMLPETVTGLTLRREGQOERLVTSIAEAWNGLPVAMTSLLPATAS 2309
QY 56 ----- 55
Db 2310 RPLGPTYAQAERYMLENTPAALATGDDWRYRIDWKRLPAEGBSERTGLSGRLAVTPED 2369
QY 56 ----- 55
Db 2370 HSAQAAVLTALVDAGAKVEVLTAGADDREALAARLTALTGDTGFTGVVSLDGLVQV 2429
QY 56 ----- KVVILTEPE- 64
Db 2430 AMVQALGDAGIKAPLWSVTQAVSVGRLDTPADPDRAMLWGLGRVVALEHFERWAGLVDL 2489
QY 65 ----- 64
Db 2490 PAQPDAAALHLVLTALSGATGEDQIAIRTTGLHARRLARAPLHGRRTPRDMQPGTCLIT 2549
QY 65 ----- GAPNITANLTS- 76
Db 2550 GGTGALGSHAARWMAHGAHELLVSRSGEQAPGATQUTAEILTASGARVTIAACDVADPH 2609
QY 77 ----- 76
Db 2610 AMRTLLDAIPETPLTAVVHTAGALDDGIVDTLTAEQVRRRAHRAKAVCASVLDLTRDL 2669
QY 77 ----- 76
Db 2670 LDAFVFSVSTLGIPOGQNYAPHNAYLDALAAARRRATGRSAVSMGPDGGMGAAGD 2729
QY 77 ----- LLSVCGWSOTI- 87
Db 2730 GVAERLRNHGVPMDPELALAALESALGRDETAITVADIDWDRFYLAYSGRPOPLVEEL 2789
QY 88 ----- 87
Db 2790 PEVRRIIDARDSATSGQGSQAQGANPLAERLAAAAPCGTEIILLGLVRAQAAAVLMRS 2849
QY 88 ----- 87

Db 2850 PEDVAADRAFKDIGFDSLAVELRNRLTRATGLOLPATLVDPHTPTLALVSLRSEFLCD 2909
QY 88 ----- NPED- 91
Db 2910 EETADARRSAALPATVGAGAGAGACTDADDPIAIVAMSCRYPGDIRSPEDLWRMLSEGG 2969
QY 92 ----- DTD- 95
Db 2970 EGITFPPTDRGWDLGLYADPDALGRAYVREGGELHDAEFDAEFFGVSPREALAMPQ 3029
QY 96 ----- 95
Db 3030 QRMILLTTSWEAFERAGIEPASLRGSSSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPS 3089
QY 96 ----- 95
Db 3090 VASGRIAYTFGLGEPATTVDTCSSSLTALHLAVRALRSGECTMALAGVAMMATPHMFV 3149
QY 96 ----- 95
Db 3150 EFSRORALAPGRSKAFSADADAGFGAABGVGLLLVERLSDARRNGHPVLAVVRGTAVNOD 3209
QY 96 ----- 95
Db 3210 GASNGLTAPNGPSOQVIRQALADARLAPGDI DAVETHGTGTSGLGDPTEAOGLOATYCKE 3269
QY 96 ----- 95
Db 3270 RPAERPLAIGSVKSNIGHTQAAAGAAGIIMVLAMRHGTLPLKTLHADESPSHVDWANSGL 3329
QY 96 ----- 95
Db 3330 ALVTEPIDWPAGTGPRAAVSSFCISGTNAHVLEQAPDAAGEVLGADPEVSETVAMA 3389
QY 96 ----- 95
Db 3390 GTAGTSEVAEGSEASEAPAPGSREASLPGHLPHVLSAKDEQSLRGQAALHAWLSEPA 3449
QY 96 ----- 95
Db 3450 DLSADGPARLDVGYTLATSRFAHRAAAVTAARDGFLDGLATLAOGGTSAHVHLDTA 3509
QY 96 ----- GHADL- 100
Db 3510 RDGTTAFLFTGQSORPGAGRELYDRHPVPARALDEICAHLDGHLLEPLLDVMPAABGSA 3569
QY 101 ----- 100
Db 3570 EAALLDETRYTOCALFALEVALPRLVESMGMRPAALLGHSVGEIAAAHVAGWPSLADAAR 3629
QY 101 ----- 100
Db 3630 LVAARGLMOELPAGGAMLAVQAAEDEIRVWLETEBYAGRLDVAAVNGPEAAVLSGDAD 3689
QY 101 ----- 100
Db 3690 AAREAEYWSGLGRTRALRVSHAFHSAHMDGMLDGFRAVLETVEFRFRPSLTVVSNVTGL 3749
QY 101 ----- 100
Db 3750 AAGPDDLCDPEYVWRHVGTVRFLDGVRLRDLGVRTCLELGPDGVLTAADGLADTPA 3809
QY 101 ----- 100
Db 3810 DSAAGSPVGSAGSADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAGHTGPDWH 3869
QY 101 ----- 100
Db 3870 ANFAGSGAHRVLDLPTYSFRDRYWLDA PAADTAVDTAGLGLGTADHPLLCVAVSLPDRDG 3929
QY 101 ----- 100

Db	3930	LLLTGRSLRTHPWADHAVLGSVLLPGAAWVELAAHAESA	SAGLRDVRRELTLEPLVLPE	3989	
Qy	101	-----	-----	100	
Db	3990	HGGVELRVTVGAPAGEPGESAGDARPVSILHSRLADAPAGT	AWSCHATGLLATDRPELP	4049	
Qy	101	-----	-----	100	
Db	4050	VAPDRAWMPPOGAEEVPLDGLYERLDGNGLAFGLPFG	LNNAVWRYEYEFADIALPATT	4109	
Qy	101	-----	-----	100	
Db	4110	NATAPATANGGSSAAAAPYGIHPALLDASLHAIAVGL	GVDEPELVRVPFHWSGVTVHAAG	4169	
Qy	101	-----	-----	100	
Db	4170	AAAARVRLASAGTDAVSLSLTDGGRPLVSVR	TLRPVTADQAAASRVGGLMHRVAVWRP	4229	
Qy	101	-----	-----	100	
Db	4230	YALASSGEODPHATSYGPTAVLGKDELKVAAL	ESAAGVEVCLYPDLAALSQDVAAGAPAP	4289	
Qy	101	-----	-----	106	
Db	4290	RTVLAPLPAGDGAEGVRGTVARTLELLOAWLA	DEHLAGTRELILLVTRGAVRDPGSGA	4349	
Qy	107	-----	-----	106	
Db	4350	DGGEDLSHAAWGLVRTAQTENPGRFGLLD	ADDASSYRTLPSVLSAGLRDEPQLALH	4409	
Qy	107	-----	-----	106	
Db	4410	DGTIRLARLASVRPETGTAAPALAPEGTVLLT	GGTGGGLGVARHVVGWVRRLLVSR	4469	
Qy	107	-----	-----	106	
Db	4470	RGTDPAGADELVHELEALGADVSAACDVAD	REALTAVLDAI PAEHPLTAVVHTAGVLSD	4529	
Qy	107	-----	-----	106	
Db	4530	GTLPSMTTDEHVLRPKVDAAFLLDELSTP	PAYDLAAAFVMFSSAAAVFGAGQGA	4589	
Qy	107	-----	-----	106	
Db	4590	NATDALAWRRRAAGLPALSLGWLWAE	TSGMTGELQADLRMRSRAGIGISDAEGIAL	4649	
Qy	107	-----	-----	106	
Db	4650	LDAALRDDRHPVLLPLRLDAAGLRDAAG	NDPAGIPALFRDVVGARTVRRP	4709	
Qy	107	-----	-----	106	
Db	4710	GTA	CTPGTGAETAATAVTLADRAATVDGP	PARQRLLEFVVVGEVAEVLGHARGH	4769
Qy	107	-----	-----	114	
Db	4770	GFLDGFDSLVELNRNLNSAGGLA	LPATLVFDHPSPAALASHLDAELPRG	ASDQDQAG	4829
Qy	115	-----	-----	114	
Db	4830	NRNGNCTTASRTAETDALLAQLTR	LEGALVLTGLSDAPGSEEVLEHL	RLSRSMVTGE	4889
Qy	115	-----	-----	114	
Db	4890	TGTGTASGAPDGCAGAE	DRPWAAGDAGGGS	EDGAGVPDPFNNASABELFGLLDQD	4949
Qy	115	-----	-----	114	
Db	4950	MSTVNEEKYLDYLRRATADL	HEARGRLEAKAGEPVAIVGMACR	LPFGV	5009
Qy	115	-----	-----	114	
Db	5010	VAGGEDAISEFPQDRGWD	VEGLYDNP	PEATCKSYAREAGFLYEAGEFDADFFG	5069

Qy	115	-----	-----	114		
Db	5070	AMDPOORLLLEASWEAFEHAGIPAA	TARCTSVGVFTGMVHYATRLTDV	PEGIEYLGT	5129	
Qy	115	-----	-----	114		
Db	5130	GNSGSVASGRVAYTILGEGPA	VTDTACSSSILVALHLAVOALRK	GEVDMALAGGVTVMST	5189	
Qy	115	-----	-----	120		
Db	5190	PSTFEVFSRQGLAPDGRSKSF	SSTADGTWSSEGVLLVERLS	DARRKGRILAVVRGT	5249	
Qy	121	-----	-----	120		
Db	5250	AVNQDASSGLTAPNGPSOOR	VIRRALADARLTTSVDV	VVEAHGTCTRLGDP	IEAOAVIA	5309
Qy	121	-----	-----	120		
Db	5310	TYQGRDGEQPLRLGSLKSN	IGHTQAAAGSVGIKMWQMRH	GVLPKTLHVEKPTDOVDW	5369	
Qy	121	-----	-----	120		
Db	5370	SAGAVELLTEAMDMWDPK	DGGLRAAVSSFGVGTNAHV	VLBEAPAAEETPASEATPAVE	5429	
Qy	121	-----	-----	120		
Db	5430	PSVGAGLVPLVSAKTPA	ALDAQIGRLAAPASQGR	TDAADPGAVARVL	AGGRAEF	5489
Qy	121	-----	-----	120		
Db	5490	VLTGQDDFAQALTAPEGL	IRGTPSDVGRVAVFPQ	CGTQWAGMGAELLDV	SKEFAAAMA	5549
Qy	121	-----	-----	120		
Db	5550	ECESALSRVYDWSLEAV	VRQAPCAPTLERVDV	VQVPTFAMVSLAKV	QWQHGVTPQAVVG	5609
Qy	121	-----	-----	120		
Db	5610	HSQGEIAAAVYAGALT	LDAAARVVTLRSKSI	AAHLAKGKGMISL	ASEATRORIENLHG	5669
Qy	121	-----	-----	120		
Db	5670	LSIAAVNGPTATV	VSGDPTQIQELAQACEAD	GVRRARIIPVDYASH	SAHVETIESELA	5729
Qy	121	-----	-----	120		
Db	5730	AGLS	PRTPEVPFSTLEGAWITEP	VLDTGTWYRNLRH	RVGFAPAVETLATDEG	5789
Qy	121	-----	-----	125		
Db	5790	SAHPVLTMTLPETV	TGLTLRREQGOERL	VTLSAEAWTNG	LTIDWAPVLP	5849
Qy	126	-----	-----	125		
Db	5850	PTYAFORRHVWLHDS	PAVQGSVQDSWRYRID	WKRLAVADASER	AGLSGRWL	5909
Qy	126	-----	-----	125		
Db	5910	EAAPVLAALSGAGAD	PVQLDVSP	GLDRORLAATLGE	ALAAAGGAVD	5969
Qy	126	-----	-----	125		
Db	5970	GHPAPTRGTGATL	TLVQALE	DAGVAAPLWC	VTHGAVSVGR	6029
Qy	126	-----	-----	126		
Db	6030	ALEHPERWGLIDL	PSDADRAALDRMT	TVLAGGTGEDQ	VAVRASGLLAR	6089
Qy	127	ACSP	TV	-----	-----	132
Db	6090	TAS	FWQADGTVLTGA	EPAAAAARRLARD	GAGHLLHTT	6149

QY	133	-----	132
Db	6150	LAGLVAELADGATATVVTCDLTDAEAAARLLACVSDAHLPLSAVLHLPPTVDSEPLAATD	6209
QY	133	-----	132
Db	6210	ADALARVVTAKATAALHLDRLLREAAAAGRPVVLFFSSVAIWGGAGGAGYAAAGTAF	6269
QY	133	-----	132
Db	6270	DALAGQHRADGPTVTSVAVSPWEGSRVTEGATGERLRLGLRPLAPATALTALDTHLGH	6329
QY	133	-----	132
Db	6330	DTAVTIADVWSSFAPGFTTARPTGLLADLPEARRALDEQOSTTAADDTVLSRELGALT	6389
QY	133	-----	132
Db	6390	AEQORRMOELVREHLAVLNLNHPSPVAVDTGRAFRDLGFSLTATVELNRNLKNATGUALPA	6449
QY	133	-----	132
Db	6450	TLVFDYPTPTLAEFLLAELGQAGAGELPVDGGVDDEPVAIVGMACRLPGCVASPED	6509
QY	133	-----	132
Db	6510	LMRLVAGGEDAISGFPODRGWDVEGLYDPPDASGRTYCRAGGFLDEAGEFDADFFGISP	6569
QY	133	-----	132
Db	6570	REALAMDPOORLLLETSSWEAVEDAGIDPTSLQGQOVGVFAGTNGPHYEPLLRNTAEDLEG	6629
QY	133	-----	132
Db	6630	YVGTGNAASIMSRVSYTLGLECPAVTVDTCSSSLVALHLAVQALRKCEGLALAGVT	6689
QY	133	-----	132
Db	6690	VMSTPTTFVFSRQRLAEDGRSKAFASADGFGPAEGVGMVLLVERLSDARRNGHRLAV	6749
QY	133	-----	132
Db	6750	VRGSVAVQDASNGLTAPNGPSQORVIRRALADARLTATADVVDVEAHGTGTRLGDPIEAQ	6809
QY	133	-----	132
Db	6810	ALIATYQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIMVQAMRHGVLPKTLHVDPRSD	6869
QY	133	-----	132
Db	6870	QIDWSAGTVELLTEAMDPKOEGLRRAAVSFGISGTNAHIVLEAPVDEADAPADEPS	6929
QY	133	-----	132
Db	6930	VGCVVPLVSAKTPAALDAQIGRLAFAFASQGRDADPGAVARVLAGGQRAQFEHRAVALG	6989
QY	133	-----	132
Db	6990	TGQDDLAALAAPEGLVRGVASGVGRVAFVFGQGTQWAGMCAELLDVSKFEFAAAWABCE	7049
QY	133	-----	132
Db	7050	AALAPYVWNSLEAVVRQAPGAPTLERVDVVPVTPAVMVSLAKVQWHGVTPQAVVGHQS	7109
QY	133	-----	132
Db	7110	GEIAAAYVAGALSDDAARVVTLRSKISGAHLAGQGMLSLALSEAAVVERLAGFDGLSV	7169
QY	133	-----	132
Db	7170	AAVNGPTATVSGDPTQIOELAQACEADGVRARIIPVDYASHSAHVETIESELADVLGL	7229
QY	133	-----	132

Db	7230	SPOTQVPPFFSTLEGAWITEPALDGGYWRNLHRVRFAPAVETLATDEGTFHFVYESAH	7289
QY	133	-----	132
Db	7290	PVLTMALPETVTGLTLRRDNGGOHRLTTSLAEWANGLTVDWASLLPTTTTHPDLPTVA	7349
QY	133	-----	137
Db	7350	FOTERYWPQDLSAAGDITSAGLGAHEPLIGAAVALADSOGCLLTGSLSLRTHPLWADH	7409
QY	138	-----	137
Db	7410	AVAGTVLLPGTAFVELAFRAGDQVGC DLVEELTLDAPLVLRPRGAVRVQLSVGASDESR	7469
QY	138	-----	137
Db	7470	RTFGLYAHPEDPAGEAETHATGVLAAARADRTAPVADPEAMPVPPGASPVVDGLYERFA	7529
QY	138	-----	137
Db	7530	ANGYCYGPLFOGVRGVHRRGDEVFADVALPAEVAGAEGRFGLHPALLDAAVQAAGARG	7589
QY	138	-----	137
Db	7590	VRGHAAAVRLERDLLYAVGATALRVRLAPAGPDTVSVAADSSGQPVFAADSLTVLPVD	7649
QY	138	-----	137
Db	7650	PAQLAAFSPTLDALHLEMTAWDGAQAALPGAVVLGGDADGLAAALRAGGTEVLSFPDL	7709
QY	138	-----	137
Db	7710	TDLVEAVDRGETPAPATVTVACPAAGPDGPEHVREAHLSLALMQWLADERFTDGRLLV	7769
QY	138	-----	137
Db	7770	VTRDAVAASGDGLRSTGQAAVMGLGRSAQTESPCRVLVLLDAGEARTAGDATAGDGLT	7829
QY	138	-----	137
Db	7830	GDATVGTSGDAALGSALATALGSGEPOLALRDGALLVRLARAAPAAADGLAAADGLA	7889
QY	138	-----	137
Db	7890	ALPLPAALWRLPEPGTDGSLSLTAAPGDAETLAPPEPLGPGQVRIAIRATGLNFRDVL	7949
QY	138	-----	137
Db	7950	ALGHYDPALMGTEGAGVWTATGVTPLAPGDRVMGLLSGAYAPVVVADARTVARMPEG	8009
QY	138	-----	137
Db	8010	WTFAGQASVPVFLTAVYALRDLADVKPGRLLVHSAAGVGMAAVOLARHWGEVHGTA	8069
QY	138	-----	137
Db	8070	SHGKWDALRALGLDDAHIASRTLDPEFAPRAASGGAGMDVVLNSLAREFVDASRLLLGP	8129
QY	138	-----	137
Db	8130	GGRFVEMKTDVRDAERVAADHPGVGYRFPDLGEAGPERIGEMLAEVIALFEDGVLRLHP	8189
QY	138	-----	137
Db	8190	VTTWDRRARDAFRHVSQARHTGKVVLTPSGDLPDGTLLTGTGTGALGGIVARHVVGW	8249
QY	138	-----	137
Db	8250	GVRLLLVSRRTDAPCAGELVHELEALGADVSVAACDVADREALTAVLDSIPAESHPLTA	8309
QY	138	-----	137

Db 8310 VHTAGVLSGDTLPSTAEDEHVLPRKVDAAFLDELSTPGYDLAAAFVMSAAAVFG 8369
Qy 138 ----- 137
Db 8370 GAGQAYAAANATLDALAWRRRTAGLPALSLGWLWLAETSGMTGSLDTRSLRARGAT 8429
Qy 138 ----- 137
Db 8430 PMSSELTSLDLDAAMRDDPALVPIALDVAALRAQQRDGLAPLLSGLTRSGRVGGAPVN 8489
Qy 138 -----ED 139
Db 8490 QRRAAAGAGBEADTDLGRLAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAPRD 8549
Qy 140 TGFDL----- 144
Db 8550 TGFDSLTAVELNRNLNAATGLRLPATLVFDHPTPGELAGHLLDELATAAGSWAEGTSG 8609
Qy 145 ----- 144
Db 8610 DTASATDQTTAAELDRLEGLASLAPAGRPPELAARLALAAALGDDDDATDLDE 8669
Qy 145 ----- 144
Db 8670 ASDDLFSFIDKELGDSDFMANNEDKRLDKRVTAELQONTRRLREIEGRTHPEVAIVG 8729
Qy 145 ----- 144
Db 8730 MACRLPGGVASPEDLWQLVAGDGAISEFPQDRGWDVEGLYDPPDASGRTYCRSGGFLH 8789
Qy 145 ----- 144
Db 8790 DAGEFADOFFGISPREALAMPQORLSLTTAWEAIESAGIDPTALKGSLGVFVGGWHTG 8849
Qy 145 ----- 144
Db 8850 YTGQTAVQSPLEGLHVSQAALGFLSGRIAYVLGTGDPALTVDTACSSSLVALHLAVQ 8909
Qy 145 -----GVTI 148
Db 8910 ALRKGECMDALAGGVTV 8926

RESULT 12
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-6

Query Match 30.4%; Score 235; DB 10; Length 11877;
Best Local Similarity 1.1%; Pred. No. 4.1;
Matches 91; Conservative 21; Mismatches 36; Indels 8329; Gaps 25;
Qy 1 AAGG----- 4
Db 450 AAGGOAVTVDPADRNAGDFDPPDRSAFGRSRSRWGGFIEDVDRFDAFFGISPREAAEM 509

Qy 5 ----- 4
Db 510 DPQRLALELWELERAGIDPSSLTGTRTGVFAGAIWDDYATLKHROGGAALTPTHTVG 569
Qy 5 ----- 4
Db 570 LHRGIIANRLSYTLGLRGPMSVVDSSQSSSLVAVHLACESLRRESELAAGGVSLNLP 629
Qy 5 ----- 4
Db 630 DSIIGASKFGLSPDGRAYTFDARANGYVRGEGGFVWLKRLSRVADGDPVLAVIRGSA 689
Qy 5 ----- 4
Db 690 VNRGAAQGMTPDAQAQEAVALREHERAGTAPADVRYVELHCTGTPVGDPIEAAALGAA 749
Qy 5 -----ILHL----- 9
Db 750 LGTRPAGQPLLVGSKVTNIGHLEGAAGIAGLIKAVLAVRGRALPASLVNTPNPAIPPE 809
Qy 10 ----- 9
Db 810 ELNLRVNTYLPWEPEHDGQRMVGVSPFGMGNTNAHVLEAPGCRGASVVESTVGS 869
Qy 10 ----- 9
Db 870 AVCGGVVPMVVSAAALDAQIERLAAFAASRDRTDGDAGVADAGAVARVLACG 929
Qy 10 -----LLVAVGPD----- 17
Db 930 RAQFEHRAVVVSGPDDLAALAAAPGLVRGVASGVGRVAFVPPGGGTOWAGMGAEILDS 989
Qy 18 --VFOA----- 21
Db 990 SAVFAAAMAECEAALSFPYVWMSLEAVVROAPGAPTLERVVDVQPVTFVWVSLARVQHH 1049
Qy 22 ----- 21
Db 1050 GVTPOAVGHSQGEIAAAVYVAGALSLLDDAARVVTLRSKSIAAHLACKGGMLSLASEDAV 1109
Qy 22 ----- 21
Db 1110 LERLAGDGLSVAANGPTATVVSGDPVQIEELARACEADGVRAARVIPVDYASHRSQVEI 1169
Qy 22 ----- 21
Db 1170 IESELAEVLAGLSQAPRVPFFSTLEGAWITEPVLDDGYWYRNLRHVRVGFAPAVETLATD 1229
Qy 22 ----- 21
Db 1230 EGFTHFEVSAHPVLTMALPGTVTGLATLRRDNGDGLVSLAEAWANGLAVDWSPLLP 1289
Qy 22 ----HOED-----TER----- 28
Db 1290 SATCHHSIDLPTVAFQTERHWLGEIALAPAGEPAVQPAVLTEAEPALDRDEQLRVIL 1349
Qy 29 ----- 28
Db 1350 DKVRAOTAQVLGYATGGQIEVDRTPREAGCTSLTGVDLNRNINAATFGRVMAPSMIFDPPT 1409
Qy 29 ----- 28
Db 1410 PEALAEQLLLHVHGEAANAPAGAPAPVAAAGAVDEPVAIVGMACRLPGSVASPEDLWRL 1469
Qy 29 ----- 28
Db 1470 VAGGDAISEFPQDRGWDVEGLYHPDPEHPTGSYVRQGGFIENVAGFADAAFFGISPREAL 1529
Qy 29 -----YVLT 32
Db 1530 AMPDQORLLLETSSWEAVEDAGIDPTSLRGRQVGVFTGMANTHEYGPSLRDGGEGLDGYLLT 1589

QY 33 N-----LNIGAEELR----- 42
Db 1590 GNTASVMSGRVSYTLGLEPALTVDTACSSSLVALHLVAQALRKGEVDMALAGGVAVMPT 1649
QY 43 ----- 42
Db 1650 PGMFVEFSQRLAGDGRSKAFASADGTSWSEGVGLLVERLSDDARRNGHQLAVVRGS 1709
QY 43 ----- 42
Db 1710 ALNQDGASNGLTAPNGPSOORVIRRALADARLTTSDDVVVEAHGTCRLGDPLEAQALLIA 1769
QY 43 ----- 42
Db 1770 TYGCRDDEQRLGLSKLSKNIGHTQAAAGVGVKVMQAMRHGLLPKTLHVDEPSDQIDW 1829
QY 43 -----DPSLG- 47
Db 1830 SAGAVELLTEAVDMPEKODGGLURRAAVSFGISGTNAHVLEBA PVVVEGASVVEFSVG 1889
QY 48 -----AQF 50
Db 1890 SAVGGVTPWVVSASAAALDAQIERLAFAFSRDRTDADAGAVDAGAVAHVLADGQAQF 1949
QY 51 ----- 50
Db 1950 EHRAVALGAGADDLVQALADPDGLIRGTASGVGRVAFVPGGTQWAGMGAELLDSAVF 2009
QY 51 -----RVHLV----- 55
Db 2010 AAAMAECEAALSPYVDWLSLEAVVRQAPGAPTLEKRVVQPVTFVAVMVSUARVMQHHGVT 2069
QY 56 ----- 55
Db 2070 QAVVCHSOGIEAAAVVAGALPLDDAARVTLRSKSTIAHLACKGGHLSLALNEDAVLERL 2129
QY 56 ----- 55
Db 2130 SDFDGLSVAANGPTATVVGDPVQIEELAQAQKADGPRARIIPVDYASHSRQVEIESE 2189
QY 56 ----- 55
Db 2190 LAQVLAGLSPOAPRVPPFFSTLEGTWITEPVLDTGYWYRNLHRVGFAPAIETLAVDEGPT 2249
QY 56 ----- 55
Db 2250 HFVEVSAHPVLTMTLPETVTGLTLRRQGGQERLVTSLAEAWVNGLPVAVNTSLLPATAS 2309
QY 56 ----- 55
Db 2310 RPGLPTYAQAERYWLENTPAALATGDDWRYRIDMKRLPAEGBERTGLSGRWLAVTPED 2369
QY 56 ----- 55
Db 2370 HSAQAAAVLTALVDGAKVEVLTAGADDDREALAARLTALTTCGDTGVVSLDLGLVPOV 2429
QY 56 -----KVVILTEPE----- 64
Db 2430 AHWQALGDAGIKAPLWSVTQGAHSVGRLDTPADPDRAMLMGLGRVVVALEHPERWAGLVDL 2489
QY 65 ----- 64
Db 2490 PAQPDAAALHLVTALSGATGEDQIAIRTTGLHARRLABAPLHGRRPTRDQPHGTVLIT 2549
QY 65 -----GAPNITANITSS----- 76
Db 2550 GGTGALGSHAARMAHGAELLLVSRSGEQAPGATQLTAEALTASGARVTIAACDVADPH 2609
QY 77 ----- 76
Db 2610 AMRTLDDAI PAETPTTAVVHTAGALDDGIVDTLTAEQVRRRAHRAKAVGASVLDELTRDLD 2669
QY 77 ----- 76

Db 2670 LDAFVLFSSVSTLGI PQGNYAPHNAYDLAARRRATGRSAVSAMGPMWDMGGAAGD 2729
QY 77 -----LLSVCCWSOTI----- 87
Db 2730 GVAERLRNHGVPMDPELAALAESALCRDETAITVADIDWDRFYLA YSSGRPOPLVEEL 2789
QY 88 ----- 87
Db 2790 PEVRRIIDARDSATSGOGSSAOGANPLAERLAAAAPGERTEILLGLVRAQAAAVLRMRS 2849
QY 88 ----- 87
Db 2850 PEDVAADRAFKDIGFDSIAGVELRNRLTRATGLOLPATLVFDHPTPLALVLSLRSEFLGD 2909
QY 88 -----NPED----- 91
Db 2910 EETADARRSAALPATVGAGAGAGAGTADDDPIAIVAMSCRYPGDIRSPEDULRMLSEGG 2969
QY 92 -----DTDP----- 95
Db 2970 EGITPFPTRDGMWLDGLDADPDALGRAYVREGGFLHDAAEFDAEFFGVSPREALAMPQ 3029
QY 96 ----- 95
Db 3030 QRMILLTTSWEAFERAGIEPASLRSGSSTGVFICLSYQDYAARVPNAPRGVEGYLLTGSTPS 3089
QY 96 ----- 95
Db 3090 VASGRIAYTFGLEGPATTVOTACSSSLTALHLAVRALRSGETMALAGVAMMATPHMVF 3149
QY 96 ----- 95
Db 3150 EFSRQALAPDGRSKAFSADADGFGAEGVGLLVERLSDDARRNGHPVLA VVRGTAVNQD 3209
QY 96 ----- 95
Db 3210 GASNGLTAPNGPSOORVIRQALADARLAPGDIDAVETHGTGTSLGDPTEAOCLOATYKGE 3269
QY 96 ----- 95
Db 3270 RPAERPLAIGSVKSNIGHTQAAAGAAGIIMVLA MRHGTLPKTLHADEPSPHVDWANSGL 3329
QY 96 ----- 95
Db 3330 ALVTEPIDWPAGTGPRRAAVSSFGISGTNAHVLEQAPDAAGEVLGADEVPEVSETVAMA 3389
QY 96 ----- 95
Db 3390 GTAGTSEVAEGSEASEAPAPGSRASLPGHLPWVLSAKDEQSLRGQAAALHJAMLSEPA 3449
QY 96 ----- 95
Db 3450 DLSADGPARLUDVGYTLATSRITAFHRAAVTAADRDGFLDGLATLAOGGTSAHVHLDTA 3509
QY 96 -----GHADL----- 100
Db 3510 RDGTATFLTCQGSORPCAGRELYDRHPVPARALDEICAHLDGHLELP LLDVWMPAAEGSA 3569
QY 101 ----- 100
Db 3570 EAALLDETRYTCALFALEVALFRLVSESGMRPAALLGHVSGETIAAAHVAGVFSLADAAR 3629
QY 101 ----- 100
Db 3630 LVAAGRMLQBELPAGGAMLAVQAAAEIRVWLETEERYAGRDLVA VNGPEAAVLGSDAD 3689
QY 101 ----- 100
Db 3690 AAREAEATWSGLGRTRALRVSHAFHSAHMDGMLDGFRAVLETVEFRPSLT VVSNVTCL 3749
QY 101 ----- 100

Db	3750	AAGPDDLCOPEYVVRHVHVGRTYRFLDGVRLRDLGVRTCLELGPVGVLTAAMAADGLADTPA	3809
Qy	101	-----	100
Db	3810	DSAAAGSPVGPAGSPADSAAGALRPRPLLALLRRKRSETETVADALGRAHAHGTGPDWH	3869
Qy	101	-----	100
Db	3870	AWFAGSGAHRVDLPTYSPRRORYWLDAPAADTAVDTAGLGLGTADHPLLGAVVSLPDRDG	3929
Qy	101	-----	100
Db	3930	LLLTGRLSLRTHPWLADHAVLGSLVLLPGCAAMVELAAHAESAAGLRDVRLELTLLPELVLPE	3989
Qy	101	-----	100
Db	3990	HGGVELRTVGAPAGEPGESAGDGPVSLHSLRLADAPAGTAWSCHATGLLATDRPELP	4049
Qy	101	-----	100
Db	4050	VAPDRAAMWPPQGAEEVPLDGLYERLDONGLAFGPLFQGLNAVWRYEGEVFADIALPATT	4109
Qy	101	-----	100
Db	4110	NATAPATANGGSAAPYGIHPALLDASLHAIAVGGIVDEPELVRVPFHWSGVTVHAAG	4169
Qy	101	-----	100
Db	4170	AAARVRLASAGTDAVSLSLTDGGRPLVSVRLTLRPVTADQAAASRVGGLMHRVAMRP	4229
Qy	101	-----	100
Db	4230	YALASSGEODPHATSYGPTAVLGKDELKVAALLESAGVEVGLYPDLAALSQDVAAAGAPAP	4289
Qy	101	-----VLYITR-----	106
Db	4290	RTVLAPLPAGPADGGAEGVGTGVTARTLELLQWLADHLAGTRLLLVTRGAVRDPESSGA	4349
Qy	107	-----	106
Db	4350	DGGEDLSHAAAGLVRTAOTENPGRFGLLDLADDASSYRTLPSVLS DAGLRDEPQALH	4409
Qy	107	-----	106
Db	4410	DGTIRLARLASVRPETGTAPALAPEGTVLLTGGTGGGLGVARHVVGWVRLLLVSR	4469
Qy	107	-----	106
Db	4470	RGTDAPGADELVHELEALGADVSAACDVADREALTAVLDALPAEHPLTAVVHTAGVLS	4529
Qy	107	-----	106
Db	4530	GTLPSMTTDEVHVRPKVDAFLDLBELTSTPAYDLAAFMFSSAAAVFGGQGYAAA	4589
Qy	107	-----	106
Db	4590	NATLDALAWRRRAAGLPALSLGWLWAEVSGTGTGELGQADLRMSRAGIGISDAEGIAL	4649
Qy	107	-----	106
Db	4650	LDAALRDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVVGARTVRRAPSAASATTA	4709
Qy	107	-----	106
Db	4710	GTACTPGTADGAETAATVTLADRAATVDPARQRLLEFVVGVEAEVLGHARGHRIDAER	4769
Qy	107	-----FDLELPDG-----	114
Db	4770	GFLDLGFDLSLTAVELNRNLNSAGGLALPATLVFDHPSPALASHLDLPRGASDQDGAG	4829
Qy	115	-----	114
Db	4830	NRNGNNGTTASRSTAETDALLAQLTRLEGALVLTGLSDAPGSEVLEHLRSLRSMVTGE	4889
Qy	115	-----	114
Db	4890	TGTGTASGAPDGAGGAEDRPMAAGDAGCGGSEDGAGVPDFMNASAEELFGLLDDQDSTD	4949
Qy	115	-----	114
Db	4950	MSTVNEEKYLDVLRATADLHEARGRLRELEAKAGEPVAIVGMACFLPGGVASPEDLWRL	5009
Qy	115	-----	114
Db	5010	VAGGEDAISEFFPODRGWDVEGLYDNPENATGKSAREAGFLYEAGEFDADFFGISPREAL	5069
Qy	115	-----	114
Db	5070	AMDPOORLLLEASWEAFEHAGIPAAATAGTSGVFTGVNMYHDYATRLTDPPEGIEYLG	5129
Qy	115	-----	114
Db	5130	GNSGSVASGRVAYTLGLEGPVTVDTACSSSLVALHLVAQLRKGEVDMALAGGVTVMST	5189
Qy	115	-----NRQ-----	120
Db	5190	PSTFVEFSRQRLAPDGRSKSPSTADGTWSSEGVLLVERLSDARRKHRLAVVRGT	5249
Qy	121	-----	120
Db	5250	AVNQDASSGLTAPNGPSOORVIRRALADARLTTSDDVVVEAHGTGTRGLDPIEAQAVTA	5309
Qy	121	-----	120
Db	5310	TYQGRDGEQPLRGLSKNSIGHTOAAAGVSGVIKWQVQMRHGVLPKTLHVEKPTDQVDM	5369
Qy	121	-----	120
Db	5370	SAGAVELLTEAMDWPKDGLRRAAVSFGVGTNAHVLEAPAAAEETPASEATPAVE	5429
Qy	121	-----	120
Db	5430	PSVGAGLVPLVSAKTAPALDAQIGRLAAFAASQGRTDADPGAVARVLAGGRAEPEHRAV	5489
Qy	121	-----	120
Db	5490	VILGTQDDFAQALTAPEGLIRGTPSDVGRVAVFPQOGTQWAGMGAELLDDVSKEFAAANA	5549
Qy	121	-----	120
Db	5550	ECESALLSRYVDWNSLEAVVRQAPGAPTLERVDVVPVTFVAVMVSLAKVMQHGVTPQAVVG	5609
Qy	121	-----	120
Db	5610	HSOGEIAAYVAGALTDDAARVTVLRSKSIAHLAGKGMISLALSEATRQRIENLHG	5669
Qy	121	-----	120
Db	5670	LSIAAVNGPTATVVSQDPTQIOELAQACEADGVRARIIPVDYASHAHVETIESELAEVL	5729
Qy	121	-----	120
Db	5730	AGLSPRTEVPFPFTLEGAWITEPVLDGTYWYRNLHRVGFAPAVETLATDEGFTHFIEV	5789
Qy	121	-----VTQLG-----	125
Db	5790	SAHPVLTMTLPETVTLGLTLRREGOGERLVTSLAEAMTNGLTIDWAPVLPATGHHPEL	5849
Qy	126	-----	125
Db	5850	PTYAFORRHVWLHDSPAVQGSVQDSWRVYRIDWKRLAVADASERAGLSGRWLVVVPEDRSA	5909
Qy	126	-----	125
Db	5910	EAAPVLAALSGAGDPVOLDVSPLDGRORLAATLGEALAAAGGAVDGVLSLLAWDESANP	5969

QY	126	-----	125
Db	5970	GHPAFFTRCTGATLTILVQALEDAGVAAPLWCVTHGAVSVGRADHVTSPAQAMVMGMRVA	6029
QY	126	-----	126
Db	6030	ALEHPERMGLIDLPSDADRAALDRMTTVLAGGTGEDOVAVRASGLLARRLVRASLPAHG	6089
QY	127	ACSPTM-----	132
Db	6090	TASPMQADGTVLVTCABEPAAAEAAARLARDGAGHLLHHTTPSGSEGAEGTSGAAEDSG	6149
QY	133	-----	132
Db	6150	LAGLVAELADLCATATVVTCDLTDAEAAARLLAGVSDAHPLSAVLHLPTVDSBPLAATD	6209
QY	133	-----	132
Db	6210	ADALARVVTKATAALHDLRLREAAAAGRPVVLVLFSSVAAIWMGAGQGAYAGTAFI	6269
QY	133	-----	132
Db	6270	DALAGOHRAAGTIVTSVAMSPWEGSRVTEGATGERLRLGLRPLAPATALTALDHALGHG	6329
QY	133	-----	132
Db	6330	DTAVTIADVDSFPAGFTTARPGTLLADLPEARRALDEQQSTTAADDTVLSRELCAITG	6389
QY	133	-----	132
Db	6390	AEQRRMOELVREHLAVLHNHPSPEAVDTGRAFRDLGFDLSLTAVELNRLKNATGLALPA	6449
QY	133	-----	132
Db	6450	TLVFDYPTERTLAFFLAEILGEQAGAEQLPVDGGVDDEPVAIVGMACRLPGGVASPED	6509
QY	133	-----	132
Db	6510	LWRLVAGGEDAISFPQDRGWDVEGLYDPDPDASGRTYCRAGGFLDEAGEFDADFFGISP	6569
QY	133	-----	132
Db	6570	REALAMDPOORULLETSMWEAVEDAGIDPTSLQQQVGVFAGTNGPHYEPLLRNTAEDLEG	6629
QY	133	-----	132
Db	6630	YVGTGNAASIMSGRVSYTLGLEGPVTVDTACSSSLVALHLAVQALRKECGLAGAGVT	6689
QY	133	-----	132
Db	6690	VNSTPTTTFVEFSRQRGLAEDGRSKAPASADGFGPAEGVGMILLVERLSDARRNGHRVLAV	6749
QY	133	-----	132
Db	6750	VRGSVNDGASNGLTAPNGPQQRVIRRALADARLTADVDVVEAHGTGTRIGDPIEAQ	6809
QY	133	-----	132
Db	6810	ALIATYGOGRDTEQPLRLGSLKSNIGHTOAAAAGVSGIIKMWQAMRHGVLPKTLHVDRPSD	6869
QY	133	-----	132
Db	6870	QIDWSAGTVELLTEAMDWPCKOEGGLRRAAVSSFGISGTNAHIVLEEAPVDEDAPEADEPS	6929
QY	133	-----	132
Db	6930	VGVVVPWLVSAKTPAALDAQIGRLAAAFASQGRTDADPGAVARVLAGGRAQFEHRAVALG	6989
QY	133	-----	132
Db	6990	TGQDDLAALAAPEGLVRGASGVGRVAFVFPQGTQWAGMGAELLDVSKFEFAAMAEC	7049
QY	133	-----	132

Db	7050	AALAPYVDWSLEAVVROAPGAPTLLERVDDVQPTFAVMVSLAKVMQHGHVTPQAVVGHSG	7109
QY	133	-----	132
Db	7110	GEIAAAYVAGALSULLDAAARVVTLRKSKISGAHLAGOGGMLSLALSEAAVVERLAGFDGLSV	7169
QY	133	-----	132
Db	7170	AAVNGPTATVVSGDPTQIQELAQACEADGVRARIIPVDYASHSAHVETIESELADVLACL	7229
QY	133	-----	132
Db	7230	SPQTPQVFFSTLEGAWITEPALDGGYWYRNLHRVRFAPAVETLATDEGFTHFVEVSAH	7289
QY	133	-----	132
Db	7290	PVLTMALPETVTGLTLLRRDNGGQHLRTTSLAEAWANGLTVDWASLLPTTTTHPDLPIYA	7349
QY	133	-----	137
Db	7350	FQTERYWPQDLSAAGDITSAGLGAABHPLLGAVALADSDGCLLTGSLSLRTHPWLADH	7409
QY	138	-----	137
Db	7410	AVAGTVLLPGTAFVELAFRAGDQVGCIDLVEELTLDAPLVLPRRGAVRVQLSVGASDESOR	7469
QY	138	-----	137
Db	7470	RTFGLYAHPEDAPGEAEWTRHATGVLAAARORTAPVADPEAMPPEAGPVDVDGLYERPA	7529
QY	138	-----	137
Db	7530	ANGYGYGPLFOGVRGVRRGDEVFADVALPAEVAGAEGRFGLHPALLDAAVQAAGACRG	7589
QY	138	-----	137
Db	7590	VRRGHAAAVRLERDLYAVGATALRVRLAPAGPDTVSVAADSSQCPVFAADSLTVLPVD	7649
QY	138	-----	137
Db	7650	PAQLAAFSDDPTLDALHLLLEWTAMDGAQAOLPGA VVLGSDADGLAALRAGTEVLSFPDL	7709
QY	138	-----	137
Db	7710	TDLVEAVDRGETPAPATVLVACPAAGPDGPEHVREALHGSALMQAWLADERFTDGLRVL	7769
QY	138	-----	137
Db	7770	VTRDAVAARSGDGLRSTGQAAVWGLGRSAQTESPCRFVLLDLAGEARTAGMATAGDGLTT	7829
QY	138	-----	137
Db	7830	GDAVGTSGDAALGSLATALGSGEPQALURDGLLVPLRLARAAAPAAADGLAAADGLA	7889
QY	138	-----	137
Db	7890	ALPLPAAPALWRLEPGTDCSLESLETAAPGDAETLAPEPLGPGQVRIATRATGLNFRDVL	7949
QY	138	-----	137
Db	7950	ALGMYPPALMGTEGAGVVVTTATPGVTHLAPGDRVMGLLSGAYAPVVVADARTVARMPEG	8009
QY	138	-----	137
Db	8010	WTFAGQASVPVFLTVYALRDLADVKPGERLLVHSAAGVGVMAAVQLARHWGEVHGTA	8069
QY	138	-----	137
Db	8070	SHGKWDALRALGLDDAHIASSRTLDFFESA FRAASGGGMDVVVLSLAREFVDASLRLGCP	8129
QY	138	-----	137

Db 8130 GGRFVEMGKTDVDRDAERVAADHPGVGYRAFDLGEAGPERIGEMLAEVIALPEDGVLRLP 8189
Qy 138 ----- 137
Db 8190 VTTWVRRRDAFRHVSQARHTGKVLVLTMPGLDPEGVLLTGTGALGGIVARHVVGW 8249
Qy 138 ----- 137
Db 8250 GVRLLVSRRTDAPGAGELVHEALGADVSAACDVADREALTAVLSDIPAELPLTA 8309
Qy 138 ----- 137
Db 8310 VHTAGVLSGDTLPSMTAEDVEHLRPKVDAFLLDELSTPGVDLAFFVMFSSAAAVFG 8369
Qy 138 ----- 137
Db 8370 GAGOCAYAAANATLDALAWRRRTAGLPALSLGWLWAEISGMTGGLSDTDRSLRARGAT 8429
Qy 138 ----- 137
Db 8430 PMDSELTLSLLDAAMRRDDPALVPALDVAALRAQORDGMLAPLLSGLTRGSRVGGAPVN 8489
Qy 138 -----ED 139
Db 8490 QRRAAAGGAGEADTLGGRLAAMPDDORVAHLRDLVRTHVATVLGHGTPSRVDLERAFRD 8549
Qy 140 TGFPL----- 144
Db 8550 TGFDSLTAVELNRNLNAATGURLPATLVFDHPTPELAGHLLDELATAAGGMAEGTSG 8609
Qy 145 ----- 144
Db 8610 DTASATQRTTAAAEULRLEGVLASLAPAGGRPELAARLALAAALGDDGDDATDLDE 8669
Qy 145 ----- 144
Db 8670 ASDDDLFSFIDKELGSDFMANNEDKLDYLRVTAELOQNTRELREIEGRTHEPVAIG 8729
Qy 145 ----- 144
Db 8730 MACRLPGVASPEDLMOLVAGDGAISEFPQDRGWDVEGLYDPDPDASGRTYCRSGFLH 8789
Qy 145 ----- 144
Db 8790 DAGEFDADFGISPREALAMPDQORLSLTTAWAIESAGIDPTALKSGGLGVFGGWHG 8849
Qy 145 ----- 144
Db 8850 YTSQTTAVQSPELEHLVSCAALGFLSGRIAYVLGTDGPALTVDTACSSSLVALHLAVQ 8909
Qy 145 -----GVTI 148
Db 8910 ALRKGECMDALAGGVTV 8926

RESULT 13
US-09-788-043C-5
; Sequence 5, Application US/09788043C
; Patent No. US20020107361A1
; GENERAL INFORMATION:
; APPLICANT: Heller, Renu
; APPLICANT: Zuo, Fengrong
; APPLICANT: Klonowski, Paul
; TITLE OF INVENTION: No. US20020107361A1el Metalloproteases Having
; TITLE OF INVENTION: Thrombospondin Domains and Nucleic Acid Compositions
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: ROCH-004
; CURRENT APPLICATION NUMBER: US/09/788,043C
; PRIOR FILING DATE: 2001-02-16
; PRIOR FILING DATE: 60/184,152
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5
; LENGTH: 1690
; TYPE: PRT
; ORGANISM: human
US-09-788-043C-5

Query Match 29.9%; Score 231; DB 10; Length 1690;
Best Local Similarity 17.9%; Pred. No. 0.071; Indels 188; Gaps 6;
Matches 60; Conservative 24; Mismatches 63; Indels 188; Gaps 6;

Qy 1 AAGGILHLEL-----LVAVG----- 15
Db 57 AGGSFSLSELPALRKRDVSRRDAPAFYELQYRGRELRFNLTAQHLLAPGCVSETRR 116
Qy 16 -----PD-----VFOAQHSD----- 25
Db 117 RGGLGRAHRAHTPACHLLGEVDPEGLGAAISACDGLKGVFOLSNEDYFIEPLDSAP 176
Qy 26 ----- 25
Db 177 ARPGHAQPHVYKQAPERLAORGSSAPSTCGVQVYPELEPRRERWEOQWRRLRR 236
Qy 26 -----TERYVLTNLNIGALLRDPSSLGAQFRVHLV 55
Db 237 LHQSVSKKQVETLVADAKMVEYHGQPOVESYVLTINMVAGLFHDFPSIGNPIHITIV 296
Qy 56 KMVLTEPEGAPNITANLTSSLLSVCGWQNTINPEDDTPGHADLVLYITRFDLELPDGN 115
Db 297 RLVLLEDEDEEDLKITHADNTPKSFCKQKQKSNMKGDAHPLHHDTAILLTRKDL-CATWN 355
Qy 116 R--QVRGVTOLGGACSPWCLITEDTGFOLGVTI 148
Db 356 RPECETGLSHVAGMCPHRSCSINEDTGLPLAFTV 390

RESULT 14
US-09-321-987B-2
; Sequence 2, Application US/09321987B
; Patent No. US20020102210A1
; GENERAL INFORMATION:
; APPLICANT: Kimble, Judith E
; APPLICANT: Bleiloch, Robert H
; TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
; FILE REFERENCE: 960296, 95386
; CURRENT APPLICATION NUMBER: US/09/321,987B
; CURRENT FILING DATE: 1999-05-28
; PRIOR FILING DATE: 60/087,170
; PRIOR FILING DATE: 1998-05-29
; PRIOR FILING DATE: 60/129,023
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2150
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-321-987B-2

Query Match 29.6%; Score 229; DB 10; Length 2150;
Best Local Similarity 3.7%; Pred. No. 0.16;
Matches 61; Conservative 20; Mismatches 65; Indels 1507; Gaps 8;

Qy 3 GGILHL----- 8
Db 5 GGSFHLQPVVAAILLVCLVYALOSGSGTISEFSSDLVFSRAKYSGVPVHHHRWODA 64
Qy 9 ----- 8
Db 65 GIHVIDSHHIVRRDSYGRGRKRDVTSTDRRRRLOQVARDGCHACHLRLSRDSDAVYVHLH 124
Qy 9 ----- 8
Db 125 RNQIIPDSHNKSVPHFSNSNFAPMVLVLDSEEVGSGMSRTDPDCIYRAHVKGVOHSIV 184

```
Qy 9 ----- 8
Db 185 NLCDSGLYGLMALPSGIHTVEPIISNGNTEHDGASRHRQHLVRFKDFPMHFKSFDHLNS 244
Qy 9 -----
Db 245 TSVNETETTVATWODOWEDVIERKARRAANSWDHYVELVADTKWYEHGRSLEDYV 304
Qy 31 LTNLNIGAEILLRDSLGAQFRVHLVKMVLITPEGAPNITANLTSSILLSVCGWSOTINPE 90
Db 305 LTLFSTVASYRHSRLASINVVVKLVILKTENAGPRITTONAQOTLQDFCRWQOYYNDP 364
Qy 91 DTDPGHADLVLYITRFDLELPDGNRQVRGVTOLG----- 125
Db 365 DSSVQHHDVAILLTRKIDICRSQKCDTLGLAELGTWCDMOKSCAIIEDNGLSAFTIAH 424
Qy 126 ----- 125
Db 425 ELGHVFSIPHDERKCYMPVKNKNPHIMAPLTLYNTHPWSWSPCSAGMLERFLENNRG 484
Qy 126 ----- 125
Db 485 OTQCLFDOPVERRYEDVFRDEPGKYYDAHQOCKFVFGPASELCYMPCTCRRLWCATFY 544
Qy 126 ----- 125
Db 545 GSQWCKRTOHPWADGTPCDESRSFMFCHGACVRLAPESLTKIDGQWDMRSWGECRTC 604
Qy 126 ----- 125
Db 605 GGGVQKGLRDCSPKPRNGGKYCVGORERYSCNTQECPMDTQPYREVOCSEFNNDIGI 664
Qy 126 -----GA 127
Db 665 QGVASTNTHVPKYANVAPNERCKLYCRLSGSAAFYLLRDVKVVDGTPCDRNGDDICVAGA 724
Qy 128 CSP----- 130
Db 725 CMPAGCDHQLHSLRRDKCGVGGDDSSCKVKVGTNEQCTFGYNEVMKIPAGSANIDIR 784
Qy 131 ----- 130
Db 785 QKYNMKNEDDNYLSRAANGEFLNGHFQVSLARQOIAFQDTVLEYSDAIIERINGT 844
Qy 131 -----TWS----- 133
Db 845 GPIRSDIYVHLVSGSHPPDISYEYMTAAVPNAVIRPISALYLMRVTTDTWTECDRACRG 904
Qy 134 -----CL-----ITED----- 139
Db 905 QQSOKLMCLDMSTHROSHDRNCQNLKPKQATRMCNIDCSTWITEDVSSCSAKCGSGOK 964
Qy 140 ----- 139
Db 965 RORVSVCKWEGDRQTPASEHLCDRNSKPSDIASCIYDCSRKNYGEWTSCTCGSNGK 1024
Qy 140 ----- 139
Db 1025 MHRKSYCVDDSNRRVDESICGREQKEATERECNRPICPRWYVGHWSECSRSCDGGVQWRH 1084
Qy 140 ----- 139
Db 1085 AOCLDAADRETHTSRCGPAQTQEHCHNEACTWQFGWSDCSAKCGDGQVQYRDANCTDRH 1144
Qy 140 ----- 139
Db 1145 RSVLPEHRLCKMEIITKCHRESCPKYKLGEMSQCSVCEDGWSRRVSCVSGNGTEVD 1204
Qy 140 ----- 139
Db 1205 MSLCGTASDRPASHQTCNLGTCPFWRNTDWSACSUSCGIGHRETRTECIYREQSVDSASF 1264
```

```
Qy 140 ----- 139
Db 1265 GDTKMPETSQTCHLLPCTSWKPSHNSPCSVTCGSGIQTRSVSCTRSGEGLIYDEYFCDRN 1324
Qy 140 -----TG----- 141
Db 1325 TRPRLKKTCEKOTCDGPRVLQKQADVPIRWATGPWTACSATCGNGTQRRLLKCRDHVR 1384
Qy 142 ----- 141
Db 1385 DLPDEYCNHLKDEVSTRNCRLCDCSYKMAEWEBCPATCGTHVQOSRNVTCVSAEDGGRT 1444
Qy 142 ----- 141
Db 1445 ILKDVDCDVQKRPSTARNCRLEPCPKGBEHWIGSWKSCASCGGMRRRSVSCTSS 1504
Qy 142 -----FD----- 143
Db 1505 SCDETRKPKMFDKNEELCPPLTNNSWQISPMWTHCSVSCGGVORRKIWCEADVLSGRKOD 1564
Qy 144 ----- 143
Db 1565 DIECSEIKPREORDCEMPPCRSYHYNKTSASMTSLSSNSNTTSSASASSLPILPPVVS 1624
Qy 144 -----LGVTI 148
Db 1625 WQTSAWSACSAKCGRGTKRRVVVECVNPSLNVTV 1657

RESULT 15
US-09-800-729-155
; Sequence 155, Application US/09800729
; Patent No. US20020068319A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: P2044P1
; CURRENT APPLICATION NUMBER: US/09/800,729
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 155
; LENGTH: 2165
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-800-729-155

Query Match 29.6%; Score 229; DB 10; Length 2165;
Best Local Similarity 3.7%; Pred. No. 0.16;
Matches 61; Conservative 20; Mismatches 65; Indels 1522; Gaps 8;

Qy 3 GGILHL----- 8
Db 5 GGSFHLLOPVVAALLLVVCLVYALQSGGTISEFSSDVLFSRAKYSVPVHHSRWQDA 64
Qy 9 ----- 8
Db 65 GIHVDSHHIVRRDSYGRGRKRDVTDTRRRRLQGVARDCGHACHLRLSDDAVIVHLH 124
Qy 9 ----- 8
Db 125 RWNQIPDSHNKSVPHFSNSFAPMVLVDSEEVGRGMSRTDPCIIYRAHVKGVQHSIV 184
Qy 9 ----- 8
Db 185 NLCDSGLYGLMALPSGIHTVEPIISNGNTEHDGASRHRQHLVRFKDFPMHFKSFDHLNS 244
Qy 9 -----ELLVAVGPDPVFOAHQEDTERYV 30
```

Db	245	TSVNETTTVATWQDQWEDVIERKARRRAANSWDHVVVADVADTKMYEYHGRSLSDYV	304
Qy	31	LTNLNIGAELLRDPSLGAQFRVHLVKMVIILTEPEGAPNITANILTSLLSYCGMSQINPE	90
Db	305	LTLFSTVASIYRHOSLRASINVVVVKLIIVLKTENAGPRITQNAQQTLODFCRWQOYYNDP	364
Qy	91	DDTPDGHADLVLYITRDELPLEDPGNRQVRGVTOLG	125
Db	365	DDSSVQHHDAVAILLTRKDIICRSQKCDTLGAEILGTMCDMQKSCAIEDNGLSAAFTIAH	424
Qy	126		125
Db	425	ELGHVFSIPHDDERKCYTPMPVANKVKFQSTKFDKTFQNNFHIMAPLEYNTHPWSWP	484
Qy	126		125
Db	485	CSAGMLERFLENNRGOTQCLFPDQPVRRYYVEDVVRDEPGKKYDAHQCKFVFGPASELC	544
Qy	126		125

Search completed: March 20, 2003, 12:40:57
Job time : 97 secs

Qy 144 ----- LGVTI 148
| | | |
| | | |

GenCore version 5.1.4.ps 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:33:39 ; Search time 22.0541 Seconds
(without alignments)
181.441 Million cell updates/sec

Title: SEQID_15

Perfect score: 716

Sequence: 1 AVGPDPVQAHQEDTERVLT.....SPTWSCLITDGTGDLGVTI 136

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	212.5	29.7	997	4	US-09-369-364A-7
2	184	25.7	1882	4	US-09-369-364A-13
3	178.5	24.9	967	4	US-09-130-491-2
4	176	24.6	859	4	US-09-369-364A-5
5	173.5	24.1	837	4	US-09-122-126B-2
6	170	23.7	245	4	US-09-369-364A-11
7	159	22.2	1205	4	US-09-491-522-11
8	157	21.9	566	4	US-09-491-522-7
9	157	21.9	1211	4	US-09-491-522-5
10	154.5	21.6	874	4	US-09-369-364A-15
11	149.5	20.9	930	4	US-09-369-364A-2
12	147.5	20.6	491	4	US-09-930-872-2
13	147.5	20.6	1224	4	US-09-930-872-4
14	147	20.5	905	4	US-09-369-364A-9
15	141.5	19.8	930	4	US-09-122-126B-15
16	133.5	18.6	1081	4	US-09-369-364A-17
17	133	18.6	608	4	US-09-130-491-13
18	72	10.1	531	2	US-08-975-114A-4
19	72	10.1	531	4	US-08-849-281A-4
20	68.5	9.6	536	2	US-08-975-114A-5
21	68.5	9.6	536	4	US-08-849-281A-3
22	66.5	9.3	273	2	US-07-857-224B-73
23	66.5	9.3	835	4	US-09-291-819-2
24	66.5	9.3	2539	4	US-09-413-814-42
25	65.5	9.1	488	4	US-08-939-309-10
26	65.5	9.1	568	4	US-08-939-309-4
27	65	9.1	315	4	US-09-134-001C-4141

28 65 9.1 793 1 US-08-015-985-3 Sequence 3, Appli
29 64.5 9.0 568 4 US-09-238-173-2 Sequence 2, Appli
30 63 8.8 739 3 US-08-510-646B-33 Sequence 33, Appl
31 62.5 8.7 737 1 US-08-453-695A-112 Sequence 112, App
32 62.5 8.7 797 1 US-08-268-161A-112 Sequence 112, App
33 62.5 8.7 797 2 US-08-453-702A-112 Sequence 112, App
34 62.5 8.7 797 4 US-09-099-639-112 Sequence 112, App
35 62.5 8.7 797 5 PCT-US95-08071-112 Sequence 112, App
36 62.5 8.7 3672 2 US-08-822-445-12 Sequence 12, Appl
37 62.5 8.7 3672 4 US-09-396-540-12 Sequence 12, Appl
38 62.5 8.7 3801 2 US-08-822-445-10 Sequence 10, Appl
39 62.5 8.7 3801 4 US-09-396-540-10 Sequence 10, Appl
40 61.5 8.6 568 4 US-08-939-309-2 Sequence 2, Appli
41 61.5 8.6 739 1 US-07-618-946B-22 Sequence 22, Appli
42 61.5 8.6 814 1 US-07-618-946B-23 Sequence 23, Appli
43 61 8.5 585 1 US-08-477-674-10 Sequence 10, Appl
44 61 8.5 585 1 US-08-473-791-10 Sequence 10, Appl
45 61 8.5 585 2 US-08-316-714-10 Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-369-364A-7
; Sequence 7, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
; US-09-369-364A-7

Query Match 29.7%; Score 212.5; DB 4; Length 997;
Best Local Similarity 37.5%; Pred. No. 6.5e-18;
Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;
Qy 11 QEDTERVLTNIGAEELLRDPGLGAQFRVHLVXNVILTEPEGAPNITANLTSSLVCG 70
Db 260 QPQVESYLTIMNVAGLPHDPSIGNPIHITIVRLVLEDEEDLKITHAQLTKSFCK 319
Qy 71 WQTINPEDTDCPADLVLYITRFDLELPGDNR--QVRGVTLGGACSPWMSCLITD 128
Db 320 WQKSINNKGDAHLPHDHTAILLRKDL-CAAMNRPCTGLSHVAGMCOPIHRSINEDT 378
Qy 129 GFDLGVTI 136
Db 379 GLPLAFTV 386

RESULT 2
US-09-369-364A-13
; Sequence 13, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31

Query Match 21.9%; Score 157; DB 4; Length 566;
Best Local Similarity 31.3%; Pred. No. 3.1e-11;
Matches 41; Conservative 24; Mismatches 60; Indels 6; Gaps 3;
QY 2 VGPVDFQAH-QEDTERVLTNLNIGAEILLRDPGLGAGFRVHLVVKVILTEPEGAPNI-TA 59
DB 274 VDDSVQFHGKEHVQKYLTLNINVIYHDESLGAINVVLRIILLSYKSKMSLIEIG 333
QY 60 NLTSSLSVCGWSTINPEDDTPGHADLVLYITRFDLEPDGNRQVRGVTVQLGGACSPT 119
DB 334 NPSQSLNVCRWAYLOOKPDTGHDYHDAIFLTRQDF----GPSMGQGYAPVTGMCHPV 389
QY 120 WSLCITEDTGF 130
DB 390 RSCTLNHEDGF 400
RESULT 9
US-09-491-522-5
; Sequence 5, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Colige, Alain
; APPLICANT: Lapierre, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B.
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1211 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-491-522-5

Query Match 21.9%; Score 157; DB 4; Length 1211;
Best Local Similarity 31.3%; Pred. No. 9.5e-11;
Matches 41; Conservative 24; Mismatches 60; Indels 6; Gaps 3;
QY 2 VGPVDFQAH-QEDTERVLTNLNIGAEILLRDPGLGAGFRVHLVVKVILTEPEGAPNI-TA 59
DB 274 VDDSVQFHGKEHVQKYLTLNINVIYHDESLGAINVVLRIILLSYKSKMSLIEIG 333
QY 60 NLTSSLSVCGWSTINPEDDTPGHADLVLYITRFDLEPDGNRQVRGVTVQLGGACSPT 119

DB 334 NPSQSLNVCRWAYLOOKPDTGHDYHDAIFLTRQDF----GPSMGQGYAPVTGMCHPV 389
QY 120 WSLCITEDTGF 130
DB 390 RSCTLNHEDGF 400
RESULT 10
US-09-369-364A-15
; Sequence 15, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 874
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-9
US-09-369-364A-15

Query Match 21.6%; Score 154.5; DB 4; Length 874;
Best Local Similarity 27.5%; Pred. No. 1.2e-10;
Matches 36; Conservative 26; Mismatches 58; Indels 11; Gaps 2;
QY 10 HQEDTERVLTNLNIGAEILLRDPGLGAGFRVHLVVKVILTEPEGAPNITANLTSSLSVC 69
DB 148 HGANLQHYILTLMSIVASYIKOSSICNLINIVNLVHNHNEOGPYINFNAOTLLKNFC 207
QY 70 GWSQTIN---PEDDTPGHADLVLYITRFDLEPDGNRQVRGVTVQLGGACSPTWSCLIT 125
DB 208 QWQHSKNYLGIGIHDIT-----AVLVTRDIEDICRAQDKCDTLGLAELGTICDPYRSCSIS 260
QY 126 EDTGFDLGVTI 136
DB 261 EDSGLSTAFTI 271

RESULT 11
US-09-369-364A-2
; Sequence 2, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 930
; TYPE: PRT
; ORGANISM: mus musculus ADAMTS-5
US-09-369-364A-2

Query Match 20.9%; Score 149.5; DB 4; Length 930;
Best Local Similarity 28.7%; Pred. No. 5.8e-10;
Matches 35; Conservative 22; Mismatches 64; Indels 1; Gaps 1;
QY 15 ERYVLNINIGAEILLRDPGLGAGFRVHLVVKVILTEPEGAPNITANLTSSLSVCWSQT 74
DB 288 QHYLLTMASIANRLYSHASIHENHRLAVVVKVVVLTOKTSLEVSKNAAATTLKNFCKWQHQ 347

Db 288 QHYLLTLASIANRLYSHASIENHIRLAVKVVVLGDKDKSLEVSXNAATTLKNFCKWQH 347
Qy 75 INPEDDTPGHADLVLYITREDLELPDGNRQVRGVTOLGACSPWCLITETGFDLGV 134
Db 348 HNQLCDDHEEHYDAAILFTREDL-CGHHSCDTLGMADVGTICSPERSCAVIEDDGLHAAP 406
Qy 135 TI 136
Db 407 TV 408

Search completed: March 20, 2003, 12:36:05
Job time : 26.0541 secs

GenCore version 5.1.4.p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:32:34 ; Search time 24.8108 Seconds
(without alignments)
526.959 Million cell updates/sec

Title: SEQID_15
Perfect score: 716
Sequence: 1 AVGPDVFQAHQEDTERVYL.....SPTWSCLITDGFGLGVTI 136

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	201	28.1	2165	T21371	hypothetical prote
2	177.5	24.8	951	T00017	gene ADAMTS-1 prot
3	172.5	24.1	837	T00355	hypothetical prote
4	159	22.2	1205	T18517	procollagen N-endo
5	112.5	15.7	860	T16892	hypothetical prote
6	85	11.9	957	T15976	hypothetical prote
7	75.5	10.5	228	C84824	hypothetical prote
8	75.5	10.5	508	T22836	hypothetical prote
9	73.5	10.3	1444	T18856	angiogenesis inhib
10	72	10.1	531	JN0586	beta-1,4-mannosyl-
11	71.5	10.0	723	T14765	hypothetical prote
12	71	9.9	236	G82052	ribulose-phosphate
13	71	9.9	443	H82957	probable glutamine
14	71	9.9	726	T08769	hypothetical prote
15	70.5	9.8	266	E71230	hypothetical prote
16	70.5	9.8	388	T36402	probable monooxyge
17	70.5	9.8	475	F70972	probable amidase -
18	70.5	9.8	539	T42624	probable translati
19	70.5	9.8	1403	T11583	probable translati
20	70	9.8	369	D81196	conserved hypotet
21	70	9.8	375	A81086	conserved hypotet
22	70	9.8	375	G81857	hypothetical prote
23	69.5	9.7	161	E83541	peptidyl-prolyl ci
24	69.5	9.7	210	T22344	hypothetical prote
25	69.5	9.7	305	D69199	conserved hypotet
26	69	9.6	157	AH3460	arsenate reductase
27	69	9.6	352	G82719	alcohol dehydrogen
28	69	9.6	875	I59350	karyopherin beta -
29	69	9.6	876	S66288	nuclear pore-targe

beta-1,4-mannosyl-
beta-1,4-mannosyl-
F8K7.10 protein -
KAP95 protein - ye
ribulose-phosphate
homeotic protein e
serine-trna ligase
importin beta chai
probable RNA helic
BIR repeat contain
conserved hypotet
CT021 hypothetical
probable multicopp
probable htpG prot
hypothetical prote
two-component sens

ALIGNMENTS

RESULT 1

T21371
hypothetical protein F25H8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21371; T24896
R:Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: 219413
A:Accession: T21371
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <WIL>
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone F25H8
R:Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: 219949
A:Accession: T24896
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <WT2>
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone T13H10
A:Gene: CESP:F25H8.3
A:Map position: 4
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/2

Query Match 28.1%; Score 201; DB 2; Length 2165;
Best Local Similarity 33.6%; Pred. No. 1.4e-12;
Matches 44; Conservative 17; Mismatches 70; Indels 0; Gaps 0;

QY 6 VFOAHQEDTERVLTNLNIGALLRDPISLGAQFRVHLVKWVLTPEGAPNITANLTSSL 65
Db 292 MYEYHGRSLEDYVLTLSFVTSIYHQSLRASINVVVVKLVLTENAGPRITONAQOTL 351
QY 66 LSVCGNSOTINPEDDTPGHADLVLYITFDLELPDGNRQVRGVGTOLGGACSPWTSCLIT 125
Db 352 QDFCRWQYNDPDSSVQHVHVAILLTRKDCRSQKCDTGLAELGTWCDMQKSCAII 411
QY 126 EDTGFDLGVTI 136
Db 412 EDNGLSAFTI 422

RESULT 2

T00017
gene ADAMTS-1 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C:Accession: T00017

hypothetical protein F57B7.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T22836

R:Lennard, N.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z19623

A:Accession: T22836

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-508 <WIL>

A:Cross-references: EMBL:Z74037; PIDN:CAA98493.1; GSPDB:GN00023; CESP:F57B7.4

A:Experimental source: Clone F57B7

C:Genetics:

A:Gene: CESP:F57B7.4

A:Map position: 5

A:Introns: 45/3; 137/2; 221/2; 256/2; 306/2; 409/3; 451/3

Query Match 10.5%; Score 75.5; DB 2; Length 508;

Best Local Similarity 29.8%; Pred. No. 5.6;

Matches 31; Conservative 15; Mismatches 55; Indels 3; Gaps 3

QY 31 DPSLGAQPRVHLVKWVILTEPEGAPNTANITLSLLSYCGWSO-TINPEDTDT-PGHADL 88

DB 192 DALISSDMPKXLRKPFVDITLSEMOENNSTMTLKIDSKAKDKFTIWLKEQTGLPRHEHA 251

QY 89 VLYITRFDELPGDNRQVRGVTLQGAGCSPTMSCLITEDTGFDL 132

DB 252 VL-ITKFDLISNGSATQGMAYVGNICENGSSSVVEDICAGL 294

RESULT 9

T18856

angiogenesis inhibitor homolog - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T18856; T24653

R:McMurray, A.

submitted to the EMBL Data Library, July 1995

A:Reference number: Z19031

A:Accession: T18856

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1444 <WIL>

A:Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GN00028; CESP:C02B4.1

A:Experimental source: Clone C02B4

R:McMurray, A.

submitted to the EMBL Data Library, July 1995

A:Reference number: Z19917

A:Accession: T24653

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1444 <WIL>

A:Cross-references: EMBL:Z50006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:C02B4.1

A:Experimental source: Clone T07C5

```
A;Experimental source: clone T07C5
A;Genetics:
A;Gene: CESP.C02B4.1
A;Map position: X
A;Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3;

Query Match      10.3%; Score 73.5; DB 2; Length 1444;
Best Local Similarity 30.6%; Pred. No. 33;
Matches 22; Conservative 11; Mismatches 34; Indels 5; Gaps 3
```


Search completed: March 20, 2003, 12:35:34
Job time : 28.8108 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:32:11 ; Search time 16.5405 Seconds
(without alignments)
341.028 Million cell updates/sec

Title: SEQID_15

Perfect score: 716

Sequence: 1 AVGPDVFOAHQEDTERVYLT.....SPTWSCLITDGTGDLGVTI 136

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	212.5	29.7	997	1 AT57_HUMAN	Q9ukp4 homo sapien
2	211.5	29.5	1593	1 AT12_HUMAN	P58397 homo sapien
3	187.5	26.2	1629	1 AT59_HUMAN	Q9p294 homo sapien
4	178.5	24.9	967	1 AT51_HUMAN	Q9uh18 homo sapien
5	177.5	24.8	967	1 AT51_RAT	Q9uhq1 rattus norv
6	177.5	24.8	968	1 AT51_MOUSE	P97857 mus musculu
7	176	24.6	860	1 AT54_HUMAN	Q9ukp5 homo sapien
8	172.5	24.1	837	1 AT54_HUMAN	Q75173 homo sapien
9	172	24.0	1077	1 AT10_HUMAN	Q9h324 homo sapien
10	170	23.7	890	1 AT58_HUMAN	Q9up79 homo sapien
11	168.5	23.5	630	1 AT54_RAT	Q9esp7 rattus norv
12	159	22.2	1205	1 AT52_BOVIN	P79331 b adamts-2
13	157	21.9	1211	1 AT52_HUMAN	Q95450 h adamts-2
14	153	21.4	1205	1 AT53_MOUSE	O15072 homo sapien
15	148.5	20.7	930	1 AT55_MOUSE	Q9r001 mus musculu
16	147	20.5	905	1 AT58_MOUSE	P57110 mus musculu
17	141.5	19.8	930	1 AT55_HUMAN	Q9una0 homo sapien
18	82	11.5	1862	1 GSAB_PICPA	Q9hfr4 pichia past
19	72	10.1	531	1 GNT3_HUMAN	Q09327 h beta-1,4-
20	70.5	9.8	475	1 AM14_MYCTU	O50404 mycobacteri
21	70.5	9.8	1403	1 YDF3_SCHPO	Q10475 schizosacch
22	70	9.8	359	1 PTN7_RAT	P49445 rattus norv
23	69	9.6	875	1 IMB1_RAT	P52296 rattus norv
24	69	9.6	876	1 IMB1_MOUSE	P70168 mus musculu
25	68.5	9.6	536	1 GNT3_MOUSE	Q10470 m beta-1,4-
26	68.5	9.6	536	1 GNT3_RAT	Q02527 r beta-1,4-
27	68.5	9.6	861	1 IMB1_YEAST	O06142 saccharomyc
28	68	9.5	372	1 HMEN_BOWMO	P27609 bombyx mori
29	68	9.5	876	1 IMB1_HUMAN	Q14974 homo sapien
30	67.5	9.4	536	1 CUEO_SALTI	Q82961 salmonella
31	67.5	9.4	536	1 CUEO_SALTY	Q82962 salmonella
32	67.5	9.4	647	1 HTPG_MYCTU	O50667 mycobacteri
33	67.5	9.4	801	1 CDB6_HUMAN	Q9un66 homo sapien

34	67	9.4	534	1	EPAL_MOUSE
35	67	9.4	575	1	VG05_BPT4
36	67	9.4	789	1	AD07_RAT
37	67	9.4	859	1	Y0D3_CABEL
38	66.5	9.3	184	1	YGB8_YEAST
39	66.5	9.3	467	1	DNPE_SCHPO
40	66	9.2	504	1	CPD9_MOUSE
41	66	9.2	1083	1	MAN1_YEAST
42	65.5	9.1	451	1	VDR_CHICK
43	65.5	9.1	616	1	RRPO_PLRV1
44	65.5	9.1	616	1	RRPO_PLRVW
45	65.5	9.1	1077	1	YBK1_YEAST

ALIGNMENTS

RESULT 1
AT57_HUMAN
ID AT57_HUMAN STANDARD; PRT; 997 AA.
AC Q9UKP4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-7 precursor (EC 3.4.24.-) (A disintegrin and metalloprotease
DE with thrombospondin motifs 7) (ADAM-TS 7) (ADAM-TS7).
GN ADAMTS7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99395124; PubMed=10464288;
RA Hurekainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, Novel Members of a New Family of
Zinc Metalloproteases";
RL J. Biol. Chem. 274:25555-25563(1999).
CC -|- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (By similarity).
CC -|- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG.
CC -|- LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS.
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -|- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
SIMILARITY).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF140675; RAD56358.1; -;
CC HSSP; P15167; IATL.
CC MEROPS; M12.231; -;
CC Genew; HGNC:223; ADAMTS7.
CC MIM; 605009; -;
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR00130; Zn_M12Bptdse.
CC Pfam; PF00090; tsp.1; -;
CC Pfam; PF01421; Reprolysin; 1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC SMART; SM00209; TSP1; 2.

```
DR PROSITE: PS00142; ZINC PROTEASE; 1.
DR PROSITE: PS0215; ADAM_MEPRO; 1.
DR PROSITE: PS0092; TSP1_1.
DR PROSITE: PS0427; DISINTEGRIN_1; FALSE NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
FT Repeat; Extracellular matrix.
FT SIGNAL 1 27
FT PROPEP 28 232 BY SIMILARITY.
FT CHAIN 233 997 ADAMTS-7.
FT SITE 204 204 CYSTEINE SWITCH (POTENTIAL).
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 389 389 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 462 537 DISINTEGRIN-LIKE.
FT DOMAIN 538 594 TSP TYPE-1 1.
FT DOMAIN 595 697 CYS-RICH.
FT DOMAIN 698 914 SPACER.
FT DOMAIN 915 990 TSP TYPE-1 2.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 997 AA; 109694 MW; 6587044ED02FC104 CRC64;

Query Match 29.7%; Score 212.5; DB 1; Length 997;
Best Local Similarity 37.5%; Pred No. 7e-14;
Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;

Qy 11 QEDTERVLTNINICAEELRDLPSLQAFRVHLVKVILTEPGANITANLTSSLLSVCVG 70
Db 260 QPQVESYVLTIMMVAGLPHDPSIGNPIHITIVRLVLEDEEDLKITHADNTUKSPCK 319

Qy 71 WQSTNPEDDTPGHADLVLYITRDELPLEDQNR--QVRGVTOLGACSPWTSCLITEDT 128
Db 320 WOKSINMGDAHPLHDDHTAILTRKDL-CAAMNRPCETGLSHVAGMCPHRSCSINEDT 378

Qy 129 GDLGVTI 136
Db 379 GLPLAFTV 386

RESULT 2
AT12 HUMAN STANDARD; PRT; 1593 AA.
AC P58397;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-
DE TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]_taxid=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RC MEDLINE=21264577; PubMed=11279086;
RA Cal S., Arqueles J.M., Fernandez P.L., Lopez-Otin C.;
RT "Identification, characterization, and intracellular processing of
RT ADAM-TS12, a novel human disintegrin, with a complex structural
RT organization involving multiple thrombospondin-1 repeats.";
RL J. Biol. Chem. 276:17932-17940(2001).
CC -!- COPACITOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
CC expressed in gastric carcinomas and in cancer cells of diverse
CC origin.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
```

```
CC SIMILARITY).
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.
CC -!- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1
CC DOMAINS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AJ250725; CAC20419.1; .
CC Genew; HGNC:14605; ADAMTS12.
CC MIM; 606184; .
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; Reptolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR000130; Zn_MTpeptdase.
CC Pfam; PF00090; tsp_1; 6.
CC Pfam; PF01421; Reptolysin; 1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC SMART; SM00209; TSP1; 8.
CC PROSITE; PS0215; ADAM_MEPRO; 1.
CC PROSITE; PS00437; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE; PS00092; TSP1; 2.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
Repeat; Extracellular matrix.
FT SIGNAL 1 25
FT PROPEP 26 240 BY SIMILARITY.
FT CHAIN 241 1593 ADAMTS-12.
FT DOMAIN 465 544 DISINTEGRIN-LIKE.
FT DOMAIN 545 596 TSP TYPE-1 1.
FT DOMAIN 597 700 CYS-RICH.
FT DOMAIN 701 826 SPACER 1.
FT DOMAIN 827 881 TSP TYPE-1 2.
FT DOMAIN 886 943 TSP TYPE-1 3.
FT DOMAIN 947 995 TSP TYPE-1 4.
FT DOMAIN 996 1315 SPACER 2.
FT DOMAIN 1316 1364 TSP TYPE-1 5.
FT DOMAIN 1367 1423 TSP TYPE-1 6.
FT DOMAIN 1426 1471 TSP TYPE-1 7.
FT DOMAIN 1471 1471 TSP TYPE-1 8.
FT DOMAIN 302 305 POLY-GLU.
FT SITE 208 208 CYSTEINE SWITCH (POTENTIAL).
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 393 393 BY SIMILARITY.
FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 215 215 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 685 685 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 790 790 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 951 951 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1104 1104 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1275 1275 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1300 1300 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1320 1320 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1503 1503 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1593 AA; 177545 MW; 07F9F48E63BD83A3 CRC64;
```



```

Query Match      24.9%; Score 178.5; DB 1; Length 967;
Best Local Similarity 28.6%; Pred. No. 2.1e-10;
Matches 37; Conservative 26; Mismatches 60; Indels 5; Gaps 2;

QY 10 HQSDTERYVLTNLNIGAEILLRDPGLGAQFRVHLVWVILTEPEGAPNITANLTSSLLSVC 69
DB 274 HGSGLKHLLTLFSAARLYKHPSIRNSVLVVVKILVIHQKGPVTSNAALTNRFC 333
QY 70 GWSOTINPEDTDPGHADLVLYITRFDELPLDGNR--QVRGVTLGGACGPTWSCLITD 127
DB 334 NWQKHNPSPDRDAEHYDTAILFTFQDL---CGSQTCDTLGMADVGVCDPSRSCSVIED 390
QY 128 TGFDLGVT 135
DB 391 DGLQAFT 398

RESULT 5
ATSL_RAT
ID ATSL_RAT STANDARD; PRT; 967 AA.
AC O9MU01; O9ER11;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
CN ADAMTS1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RX [1]_TaxID=10116;
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RA Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,
RA Little S.P.;
RT "Induction of a disintegrin and metalloproteinase with the
RT thrombospondin type 1 motif (ADAMTS)";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 18-967 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=20304099; PubMed=10847486;
RA Diamantis I., Luehli M., Hoesli M., Reichen J.;
RA "Cloning of the rat ADAMTS-1 gene and its down regulation in
RA endothelial cells in cirrhotic rats.";
RT Liver 20:165-172(2000).
CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER, HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1683-GLU-1-LEU-1684
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COPACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM
CC CIRRHOTIC LIVER.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.iab-sib.ch/announce/
or send an email to license@iab-sib.ch).
-----
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.iab-sib.ch/announce/
CC CC or send an email to license@iab-sib.ch).
CC CC -----
DR EMBL; AF149118; AAD34012.1; -;
DR MEROPS; M12.222; -;
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep M12B.propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTPeptide.
DR Pfam; PF00090; tsep_1; 6.
DR Pfam; PF01421; Reprolysin; 2.
DR Pfam; PF01562; Pep M12B.propep; 2.
DR SMART; SM00209; TSP1_3;
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00442; ZINC_PROTEASE; 1.
DR PROSITE; PS00092; TSP1; 2.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 54
FT PROPEP 55 252
FT CHAIN 253 967
FT SITE 205 205
FT METAL 401 401
FT ACT_SITE 402 402
FT METAL 405 405
FT METAL 411 411
FT METAL 476 558
FT DOMAIN 559 615
FT DOMAIN 616 724
FT DOMAIN 725 857
FT DOMAIN 858 907
FT DOMAIN 908 967
FT DOMAIN 134 198
FT CARBOHYD 547 547
FT CARBOHYD 720 720
FT CARBOHYD 764 764
FT CARBOHYD 782 782
FT CARBOHYD 945 945
FT CONFLICT 21 21
FT CONFLICT 26 31
FT CONFLICT 49 49
FT CONFLICT 72 72
FT CONFLICT 79 79
FT CONFLICT 249 249
FT CONFLICT 262 265
FT CONFLICT 607 607
FT CONFLICT 936 936
FT CONFLICT 962 962
SQ SEQUENCE 967 AA; 105705 MW; F93C864F6DCDB4CF CRC64;

Query Match      24.8%; Score 177.5; DB 1; Length 967;
Best Local Similarity 28.6%; Pred. No. 2.6e-10;
Matches 36; Conservative 23; Mismatches 66; Indels 1; Gaps 1;

QY 10 HQSDTERYVLTNLNIGAEILLRDPGLGAQFRVHLVWVILTEPEGAPNITANLTSSLLSVC 69
DB 274 HGSGLKHLLTLFSAARLYKHPSIRNSVLVVVKILVIHQKGPVTSNAALTNRFC 333
QY 70 GWSOTINPEDTDPGHADLVLYITRFDELPLDGNRQVRGVTLGGACGPTWSCLITD 129
DB 334 NWQKHNPSPDRDAEHYDTAILFTFQDL---CGSHTCDTLGMADVGVCDPSRSCSVIED 392
QY 130 FDLGVT 135
DB 393 LQAFT 398

RESULT 6
ATSL_MOUSE

```


FT	DOMAIN	662	665	POLY-GLY.
FT	CARBOHYD	99	99	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	172	172	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	222	222	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	234	234	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	676	676	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	843	843	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQL	SEQUENCE	860 AA;	97098 MW;	E57213015DEC82C5 CRC64;

Query Match 24.6%; Score 176; DB 1; Length 860;
Best Local Similarity 33.8%; Pred. No. 3.3e-10;
Matches 47; Conservative 23; Mismatches 51; Indels 18; Gaps

QY	11	QEDTERVLTNLNIGIELLRDPSLGQAFRVHLVKVMLITPECAPNITAN--LTSSLSLV 68
DB	268	RKDIERYILSMVNIIVAKLYRDSLSGNVNIIVARLIVLTEDQ--PNLEINHHDKSLDSF 325
QY	69	CGWSOTI--NPEDDDTDP-----GHADLVLYITRFDL-----ELPDGNRQVRGVLTQLGAC 117
DB	326	CKVQKSLSHQSQONTIPENGIAHHDNAVLTITDYCTYKPKCG---TLGLASVAGMCE 382
QY	118	PTWSCLITDGTGDLGVTI 136
DB	383	PERSCSINEDIGLSAFTI 401

RESULT 8

ID	ATSA4_HUMAN	STANDARD;	PRT;	837 AA.
AC	Q75173; Q9UN83;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1) (ADMP-1).			
GN	ADAMTS4 OR KIAA0688.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RP	MEDLINE=98403890; PubMed=9734811;			
RX	Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,			
RA	Kotani H., Nomura N., Ohara O.;			
RA	"Prediction of the coding sequences of unidentified human genes. X.			
RT	The complete sequences of 100 new cDNA clones from brain which can			
RT	code for large proteins in vitro."			
RL	DNA Res. 5:169-176(1998).			
RN	[2]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE=99286303; PubMed=10356395;			
RA	Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,			
RA	Liu R.-Q., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,			
RA	Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,			
RA	Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,			
RA	Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,			
RA	Tzaskos J.M., Arner E.C.;			
RT	"Purification and cloning of aggrecanase-1: a member of the ADAMTS			
RT	family of proteins."			
RL	Science 284:1664-1666(1999).			
RN	[3]			
RN	SEQUENCE FROM N.A.			
RA	Sawaji Y., Nagase H., Saklatvala J., Clark A.R.;			
RT	"ADAMTS-4 genomic locus."			
RT	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RN	PARTIAL SEQUENCE, AND CHARACTERIZATION.			
RX	MEDLINE=20400518; PubMed=10827174;			
RA	Tortorella M.D., Pratta M.A., Liu R.-Q., Abbaszade I., Ross H.,			
RA	Burn T.C., Arner E.C.;			

RT "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
 RT aggrecan substrate recognition and cleavage.",
 RL J. Biol. Chem. 275:25791-25797(2000).
 CC -1- INVOLUTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
 CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
 CC DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. COULD ALSO BE A ONE
 CC CRITICAL FACTOR IN THE EXACERBATION OF NEURODEGENERATION IN
 CC ALZHEIMER'S DISEASE.
 CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393
 CC site.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED
 CC AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.
 CC -1- INDUCTION: BY INTERLEUKIN-1.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
 CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO AS ADAMTS2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AB014588; BAAJ1663.1; -
 CC EMBL: AF148213; AAD41494.1; -
 CC EMBL: AY044847; AAL02262.1; -
 CC MEROPS: M12.221; -
 CC Genew: HGNC:220; ADAMTS4.
 CC MIM: 603876; -
 CC InterPro: IPR001762; Disintegrin.
 CC InterPro: IPR001590; Reprolysin.
 CC InterPro: IPR000884; TSPI.
 CC InterPro: IPR000130; Zn_M12peptidase.
 CC Pfam: PF000090; tsp_1; 1.
 CC Pfam: PF01421; Reprolysin; 1.
 CC SMART: SM00209; TSPI; 1.
 CC PROSITE: PS00142; ZINC_PROTEASE; 1.
 CC PROSITE: PS50215; ADAM_MEPRO; 1.
 CC PROSITE: PS50092; TSPI; 1.
 CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
 CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW Extracellular matrix.
 FT SIGNAL 1 51
 FT PROPEP 52 212 POTENTIAL.
 FT CHAIN 213 837
 FT SITE 194 194 ADAMTS-4.
 FT METAL 361 361 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 362 362 BY SIMILARITY.
 FT METAL 365 365 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 371 371 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 437 519 DISINTEGRIN-LIKE.
 FT DOMAIN 520 576 TSP TYPE-1.
 FT DOMAIN 577 685 CYS-RICH.
 FT DOMAIN 686 837 SPACER.
 FT DOMAIN 247 252 POLY-ALA.
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 77 77 A -> T (IN REF. 1).
 FT CONFLICT 626 626 R -> Q (IN REF. 3).
 FT CONFLICT 682 682 G -> R (IN REF. 3).
 FT SEQUENCE 837 AA; 90224 MW; 5DF9C9AC137DF41F CRC64;
 Query Match 24.1%; Score 172.5; DB 1; Length 837;
 Best Local Similarity 30.2%; Pred. No. 7.2e-10;
 Matches 38; Conservative 20; Mismatches 67; Indels 1; Gaps 1;

QY 10 HQEDTERYVLTNLNIGAEILLRDPISGLAQPRVHLVKWVILTEPEGAPNITANLTSSLLSVC 69
 DB 234 HGAGLKRYLLTVNAAAKAFKPSINPVLSLVVTRVLVLGSGEGQVGPSSAAQTLSFC 293
 QY 70 GMSQTINPEDDDPGHADLVLYITRFDLPLDGNRQVRGVTQLGACSGPTMSCLITDTG 129
 DB 294 AMQGLNTPEDSDPDHFDTAILFTRODL-CGVSTCDTLGMADVTVCDPARSCAIVEDDG 352
 QY 130 FDLGVT 135
 DB 353 LOSAFT 358
 RESULT 9
 AT10_HUMAN STANDARD; PRT; 1077 AA.
 ID AT10_HUMAN
 AC Q9H324;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with thrombospondin motifs 10) (ADAM-TS10) (Fragment).
 GN ADAMTS10.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Apte S.S.;
 RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
 RT thrombospondin type I repeats.",
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (By
 CC similarity).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF163762; AAG35563.1; -
 CC MEROPS: M12.235; -
 CC Genew: HGNC:13201; ADAMTS10.
 CC InterPro: IPR001762; Disintegrin.
 CC InterPro: IPR002870; Pep_M12B_propep.
 CC InterPro: IPR001590; Reprolysin.
 CC InterPro: IPR000884; TSPI.
 CC InterPro: IPR000130; Zn_M12peptidase.
 CC Pfam: PF000090; tsp_1; 5.
 CC Pfam: PF01421; Reprolysin; 1.
 CC Pfam: PF01562; Pep_M12B_propep; 1.
 CC SMART: SM00209; TSPI; 5.
 CC PROSITE: PS50215; ADAM_MEPRO; 1.
 CC PROSITE: PS50092; TSPI; 2.
 CC PROSITE: PS00142; ZINC_PROTEASE; 1.
 CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
 CC Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
 KW Repeat; Extracellular matrix.
 FT NON_TER 1 1
 FT PROPEP <1 207 BY SIMILARITY.
 FT CHAIN 208 1077 ADAMTS-10.

FT METAL 366 366 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT SITE 367 367 BY SIMILARITY.
 FT METAL 370 370 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 376 376 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 434 520 DISINTEGRIN-LIKE.
 FT DOMAIN 578 679 CYS-RICH.
 FT DOMAIN 680 802 SPACER.
 FT DOMAIN 521 577 TSP TYPE-1 1.
 FT DOMAIN 799 860 TSP TYPE-1 2.
 FT DOMAIN 862 918 TSP TYPE-1 3.
 FT DOMAIN 922 976 TSP TYPE-1 4.
 FT DOMAIN 981 1031 TSP TYPE-1 5.
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 714 714 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 769 769 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 866 866 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1077 AA; 118072 MW; 3914DE18DCBFF587 CRC64;
 Query Match 24.0%; Score 172; DB 1; Length 1077;
 Best Local Similarity 31.4%; Pred. No. 1.1e-09;
 Matches 43; Conservative 24; Mismatches 56; Indels 14; Gaps 3;
 QY 11 QEDTERVLTNLNIGAEILLRDPGLGQFRVHLVKNVILTEPEGAPNITANLTSLLSVCG 70
 DB 231 RRDVEQVLAIMVIAKLVFOSSLGSTVNLVRLILLTDEOPTLEITHAGKSLDSFCK 290
 QY 71 WSOITL-----NPEDTDFCHADLVLYITRFDL-----ELPDGNRQVRGVTLGGACSP 119
 DB 291 WKSIVNHSNGHNAIPENGVAHDTAVLITRYDICIYKNKPCG---TLGLAPVGMGCERE 347
 QY 120 WSLCLTDTGDFGLGVTI 136
 DB 348 RSCSVNEDIGLPQFTI 364
 RESULT 10
 AT58_HUMAN STANDARD; PRT; 890 AA.
 AC Q9UP79; Q9NZ50;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2)
 DE (METH-8).
 GN ADAMTS8 OR METH2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=99367466; PubMed=10438512;
 RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
 Lombard M., Iruela-Arispe M.L.;
 RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
 family of proteins with angio-inhibitory activity.";
 RL J. Biol. Chem. 274:23349-23357(1999).
 RN [2]
 RP SEQUENCE OF 195-440 FROM N.A.
 RX MEDLINE=20079168; PubMed=10610729;
 RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
 RT "ADAM-TS8, a novel metalloproteinase of the ADAM-TS family located on
 mouse chromosome 9 and human chromosome 11.";
 RL Genomics 62:312-315(1999).
 CC -1- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 matrix (By similarity).
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER

CC CC EXPRESSION IN BRAIN, PLACENTA, HEART, STOMACH AND FETAL BRAIN AND
 CC CC KIDNEY.
 CC CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC CC -1- PTH: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
 CC CC SIMILARITY).
 CC CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC CC the European Bioinformatics Institute. There are no restrictions on its
 CC CC use by non-profit institutions as long as its content is in no way
 CC CC modified and this statement is not removed. Usage by and for commercial
 CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC CC or send an email to license@isb-sib.ch).
 CC CC -----
 CC CC EMBL; AF060153; AAD48081.1; -;
 CC CC EMBL; AF175283; AAF25806.1; -;
 CC CC HSSP; P34179; 1IAG.
 CC CC MEROPS; M12.226; -;
 CC CC Genew; HGNC:224; ADAMTS8.
 CC CC MIM; 605175; -;
 CC CC InterPro; IPR001762; Disintegrin.
 CC CC InterPro; IPR002870; Pep M12B_propep.
 CC CC InterPro; IPR001590; Reprolysin.
 CC CC InterPro; IPR000984; TSP1.
 CC CC InterPro; IPR000130; Zn_MTPeptdse.
 CC CC Pfam; PF00090; tsp1; 2.
 CC CC Pfam; PF01421; Reprolysin; 1.
 CC CC Pfam; PF01562; Pep M12B_propep; 1.
 CC CC SMART; SM00209; TSP1_2.
 CC CC PROSITE; PS50215; ADAM_MEPRO; 1.
 CC CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 CC CC PROSITE; PS00427; DISINTEGRIN_1; Signal; Glycoprotein; Zymogen;
 CC CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 CC CC Repeat; Extracellular matrix; Heparin-binding.
 CC CC SIGNAL 1 27
 CC CC PROPEP 28 214 BY SIMILARITY.
 CC CC CHAIN 215 890 ADAMTS-8.
 CC CC METAL 364 364 ZINC (CATALYTIC) (BY SIMILARITY).
 CC CC ACT SITE 365 365 BY SIMILARITY.
 CC CC METAL 368 368 ZINC (CATALYTIC) (BY SIMILARITY).
 CC CC METAL 374 374 ZINC (CATALYTIC) (BY SIMILARITY).
 CC CC DOMAIN 439 526 DISINTEGRIN-LIKE.
 CC CC DOMAIN 527 583 TSP TYPE-1 1.
 CC CC DOMAIN 584 690 CYS-RICH.
 CC CC DOMAIN 691 832 SPACER.
 CC CC DOMAIN 833 890 TSP TYPE-1 2.
 CC CC DOMAIN 202 205 POLY-PRO.
 CC CC CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CC CARBOHYD 401 401 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CC CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CC CARBOHYD 491 491 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CC CARBOHYD 600 600 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CC CONFLICT 195 195 E -> R (IN REF. 2).
 CC CC CONFLICT 413 410 YLTLLDGGHGLCLLDGCAALPLPTGL -> FSGCHLQGW
 CC CC SEQUENCE 890 AA; 96671 MW; 57D70EE03D5739D3 CRC64;
 CC CC IFKYLCKVSEKGLDMP (IN REF. 2).
 Query Match 23.7%; Score 170; DB 1; Length 890;
 Best Local Similarity 26.6%; Pred. No. 1.4e-09;
 Matches 33; Conservative 28; Mismatches 63; Indels 0; Gaps 0;
 QY 13 DTERVLTNLNIGAEILLRDPGLGQFRVHLVKNVILTEPEGAPNITANLTSLLSVCGWS 72
 DB 239 DLQNHILTMVAARIYKHPISKNINLMVVKVILVEDEKMGVSDNGGLTLRNFNCNQ 298
 QY 73 QTINPEDDTPGHADLVLYITRFDLPELPGNRQVRGVTLGGACSPWLSCLITDGTGDL 132
 DB 299 RRENQPSDRHPEHYDTAILLTRQNFQCGEGCLDTLGVADIGTICDPNKSVCVIEDEGLQA 358

CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
 CC COLLAGEN TYPE XIV.
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -!- TISSUE SPECIFICITY: ENZYMACTIC ACTIVITY IS DETECTED AT HIGH LEVEL
 CC IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS
 CC AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS
 CC WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -!- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF DERMATOSPARAXIS, A
 CC RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN
 CC FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN
 CC INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
 CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X96389; CAA65253.1; -;
 CC MEROPS: M12.301; -;
 CC InterPro: IPR001762; Disintegrin.
 CC InterPro: IPR002870; Pep_M12B_propep.
 CC InterPro: IPR001590; RepPolysin.
 CC InterPro: IPR000884; TSP1.
 CC InterPro: IPR000130; Zn_MTpeptdase.
 CC Pfam: PF00090; tsp_1; 4;
 CC Pfam: PF01421; RepPolysin; 1.
 CC Pfam: PF01562; Pep_M12B_propep; 1.
 CC SMART: SM00209; TSP1; 4;
 CC PROSITE: PS00215; ADAM_MEPRO; 1.
 CC PROSITE: PS00092; TSP1; 1.
 CC PROSITE: PS00142; ZINC_PROTEASE; FALSE_NEG.
 CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
 CC KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 CC Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.
 CC FT SIGNAL 1 28 POTENTIAL.
 CC FT PROPEP 29 253 BY SIMILARITY.
 CC FT CHAIN 254 1205 ADAMTS-2
 CC FT METAL 402 402 ZINC (CATALYTIC) (POTENTIAL).
 CC FT ACT_SITE 403 403 BY SIMILARITY.
 CC FT METAL 406 406 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT DOMAIN 474 554 DISINTEGRIN-LIKE.
 CC FT DOMAIN 555 611 TSP TYPE-1 1.
 CC FT DOMAIN 612 716 CYS-RICH.
 CC FT DOMAIN 717 845 SPACER.
 CC FT DOMAIN 846 905 TSP TYPE-1 2.
 CC FT DOMAIN 906 968 TSP TYPE-1 3.
 CC FT DOMAIN 969 1024 TSP TYPE-1 4.
 CC FT SITE 685 687 CELL ATTACHMENT SITE (POTENTIAL).
 CC FT DOMAIN 31 35 POLY-ALA.
 CC FT DOMAIN 177 180 POLY-GLU.
 CC FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 942 942 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 943 943 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 987 987 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 1025 1025 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 1092 1092 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1144 1144 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1205 AA; 133887 MW; 7B5B232A45320371 CRC64;
 Query Match 22.2%; Score 159; DB 1; Length 1205;
 Best Local Similarity 33.3%; Pred. No. 2.7e-08;
 Matches 44; Conservative 23; Mismatches 57; Indels 8; Gaps 5;
 QY 2 VGPDIYQAH-QEDTERYVLTNNIGALLRDPGLAQFRVHLVQVILTEPGAPNI-TA 59
 DB 268 VDDSVVQFHGTGHEVQKYLTLNNIVNEIYHDSLGAIHNVLVRIILLSYGKSMSTIEIG 327
 QY 60 NLTSSLLSVCGRNSQ-TINPEDDTDPCHADLVLYITRFDELDPGNRQVRGVLTOLGGACSP 118
 DB 328 NPSQSLSEYCRWAYLQOKPDTDHDEYH-DHAIFLTRQDF-----GPSGMOGYAPVTGWCHP 382
 QY 119 TWSCLITEDTGF 130
 DB 383 VRSCTLNHEDGF 394
 RESULT 13
 ATSS2_HUMAN
 ID ATSS2_HUMAN STANDARD; PRT; 1211 AA.
 AC Q95450;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and
 DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)
 DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
 DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI)
 DE (Procollagen I/II amino-propeptide processing enzyme).
 GN ADAMTS2 OR PCINP OR PCPNI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI), AND DISEASE.
 RC TISSUE=Skin;
 RX MEDLINE=99347935; PubMed=10417273;
 RA Collige A., Sieron A.B., Li S.-W., Schwarze U., Petty E.,
 RA Wartecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,
 RA Byers P.H., Lapiere C.M., Prockop D.J., Nussgens B.V.;
 RT "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis
 RT are caused by mutations in the procollagen I N-proteinase gene";
 RL Am. J. Hum. Genet. 65:308-317(1999).
 CC -!- FUNCTION: Cleaves the propeptides of type I and II collagen prior
 CC to fibril assembly. Does not act on type III collagen. May also
 CC play a role in development that is independent of its role in
 CC collagen biosynthesis.
 CC -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
 CC alpha-1(I) at Pro-Gln and of alpha-1(II) and alpha-2(I) chains
 CC at Ala-Gln.
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
 CC COLLAGEN TYPE XIV (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-
 CC PROCOLLAGEN PEPTIDASE ACTIVITY.
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON
 CC AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -!- DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos
 CC syndrome type VIIC (EDS-VIIC), a recessively inherited connective-
 CC tissue disorder characterized clinically by severe skin fragility
 CC and joint hypermobility and biochemically by the presence in skin
 CC of procollagen incompletely processed at the N-terminus.

CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -|- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
 CC -|- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ003125; CAA05880.1; -
 CC DR MEROPS: M12.301; -
 CC DR Genew: HGNC:218; ADAMTS2.
 CC DR MIM: 604539; -
 CC DR MIM: 225410; -
 CC DR InterPro: IPR001762; Disintegrin.
 CC DR InterPro: IPR002870; Pep_M12B_propep.
 CC DR InterPro: IPR001590; Reptolysin.
 CC DR InterPro: IPR000884; TSP1.
 CC DR InterPro: IPR000130; Zn_Mtpeptdse.
 CC DR Pfam: PF00090; tsp.1; 4.
 CC DR Pfam: PF01421; Reptolysin.1.
 CC DR Pfam: PF01562; Pep_M12B_propep; 1.
 CC DR SMART: SMO0209; TSP1; 4.
 CC DR PROSITE: PS0215; ADAM_MEROP; 1.
 CC DR PROSITE: PS0092; TSP1; 1.
 CC DR PROSITE: PS00142; ZINC_PROTEASE; FALSE NEG.
 CC DR PROSITE: PS00427; DISINTEGRIN.1; FALSE NEG.
 CC DR Hydrolase: Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 CC Repeat: Collagen degradation; Extracellular matrix; Heparin-binding;
 CC Alternative splicing; Ehlers-Danlos syndrome.
 CC SIGNAL 1 29 POTENTIAL.
 CC FT PROPEP 30 253 BY SIMILARITY.
 CC FT CHAIN 254 1211 ADAMTS-2.
 CC FT METAL 408 409 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT ACT_SITE 409 409 BY SIMILARITY.
 CC FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT METAL 418 418 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT DOMAIN 480 560 DISINTEGRIN-LIKE.
 CC FT DOMAIN 561 617 TSP TYPE-1 1.
 CC FT DOMAIN 618 722 CYS-RICH.
 CC FT SITE 691 693 CELL ATTACHMENT SITE (POTENTIAL).
 CC FT DOMAIN 723 851 SPACER.
 CC FT DOMAIN 852 911 TSP TYPE-1 2.
 CC FT DOMAIN 912 974 TSP TYPE-1 3.
 CC FT DOMAIN 975 1030 TSP TYPE-1 4.
 CC FT DOMAIN 40 43 POLY-ALA.
 CC FT DOMAIN 185 188 POLY-GLU.
 CC FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 993 993 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 1031 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 1098 1098 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 1145 1145 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 1150 1150 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT VARSPPLIC 544 566 HCFKGHCILTPDLKRDGSGA -> FRPGAVAHACVPS
 CC FT VARSPPLIC 567 1211 TLGGGRWIA (IN ISOFORM SPNP1).
 CC FT VARSPPLIC 1211 AA; 134722 MW; BECEE25C3CAD2D CRC64;
 CC SQ SEQUENCE 1211 AA; 134722 MW; BECEE25C3CAD2D CRC64;
 CC -----
 CC Query Match 21.9%; Score 157; DB 1; Length 1211;
 CC Best Local Similarity 31.3%; Pred. No. 4.4e-08;
 CC Matches 41; Conservative 24; Mismatches 60; Indels 6; Gaps 3;
 CC -----
 CC Qy 2 VGPVVFQAH-QEDTRYVLTNUNIGAEILLRDPSSLAQAFRVHLVQKVVILTEPEGAPNI-TA 59
 CC Db 274 VDDSVVQPHGKHEVKYLLTLNMVINEIYVHDSGLAHINVLVRLISYKGSMSLIBIG 333
 CC Qy 60 NLTSLLSVCVGSQINPEDDTPDGHADLVLYITRFDLELPDGNRQVRGVTLGGACSP 119

Db 334 NPSQSLNVCWAYLQKQPDTHGDEYHDHAIFLTRQDF-----GPSGNGQYAPVTGMCHPV 389
 Qy 120 WSLCLITDTGF 130
 Db 390 RSCLTNHEDGF 400
 RESULT 14
 AT33_HUMAN STANDARD; PRT; 1205 AA.
 ID AT33_HUMAN
 AC O15072; Q9BXZ8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-3 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with thrombospondin motifs) (ADAM-TS 3) (ADAM-TS3) (Procollagen II
 DE amino-proteptide processing enzyme) (Procollagen II N-proteinase) (PC
 DE II-NP).
 DE ADAMTS3 OR KIAA0366.
 GN Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE OF 1-227 FROM N.A.
 RX MEDLINE=21402912; PubMed=11408482;
 RA Fernandes R.J., Hirohata S., Engle J.M., Colige A., Cohn D.H.,
 RA Byrne D.R., Apte S.S.;
 RT "Procollagen II amino propeptide processing by ADAMTS-3. Insights on
 RT dermatoparaxis.";
 RL J. Biol. Chem. 276:31502-31509(2001).
 RL [2]
 RP SEQUENCE OF 5-1205 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RL code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 CC -|- FUNCTION: Cleaves the propeptides of type II collagen prior to
 CC fibril assembly. Does not act on types I and III collagens.
 CC -|- COPACITOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -|- TISSUE SPECIFICITY: Found in cartilage and skin.
 CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -|- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -|- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
 CC -|- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF247668; AAK28400.1; -
 CC DR EMBL: AB002364; BAA20821.1; -
 CC DR MEROPS: M12.220; -
 CC DR Genew: HGNC:219; ADAMTS3.
 CC DR MIM: 605011; -
 CC DR InterPro: IPR001762; Disintegrin.
 CC DR InterPro: IPR002870; Pep_M12B_propep.
 CC DR InterPro: IPR001590; Reptolysin.

```

DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn MTpeptdse.
DR Pfam; PF00090; tsp 1; 4.
DR Pfam; PF01421; RepPolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 4.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSP1; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 20
FT PROPEP 21 249
FT CHAIN 250 1205
FT METAL 398 398
FT ACT_SITE 399 399
FT METAL 402 402
FT METAL 408 408
FT DOMAIN 470 550
FT DOMAIN 551 607
FT DOMAIN 608 712
FT DOMAIN 713 844
FT DOMAIN 845 902
FT DOMAIN 903 965
FT DOMAIN 966 1017
FT DOMAIN 1017 1017
FT CARBOHYD 83 83
FT CARBOHYD 119 119
FT CARBOHYD 242 242
FT CARBOHYD 345 345
FT CARBOHYD 475 475
FT CARBOHYD 814 814
FT CARBOHYD 942 942
SQ SEQUENCE 1205 AA; 135570 MW; EB07B286FC95FB7 CRC64;

Query Match 21.4%; Score 153; DB 1; Length 1205;
Best Local Similarity 31.3%; Pred. No. 1.1e-07;
Matches 41; Conservative 21; Mismatches 63; Indels 6; Gaps 3;

QY 2 VGPDPQAH-OEDTERYVLTNLNIGALLRDPGLQAFRVHLVQWVILTEPGAPNI-TA 59
DB 264 VDDSVVRFPGKEHVQVYLLTLMNIVNEIYHDESGLVHINVLVRMTMLGVAKSISLIERG 323
QY 60 NLTSSLLSCGKSQTINPEDDTPGHADLVLYITREDFLELPDGNRVRGVTOLGACSP 119
DB 324 NPSRSLNVCWRWASQQQRSDLNHSEHHDHAFILTRQDF---GPAGMQGVAPVTGMCHPV 379
QY 120 WSLCLTEDTGF 130
DB 380 RSCTLNHEDGF 390

RESULT 15
AT55 MOUSE STANDARD; PRT; 930 AA.
AC O9R061.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)
DE (ADMP-2) (Implantin).
GN ADAMTS5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99395124; PubMed=10464288;
RA Hurekainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of

```

```

zinc metalloproteases.";
J. Biol. Chem. 274:25555-25563(1999).
-|- INVOLVED IN CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
-|- FUNCTIONED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
-|- DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
-|- PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
-|- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393
site.
-|- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
-|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (By similarity).
-|- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PERI-
-|- IMPLANTATION PERIOD IN EMBRYO AND TROPHOBLAST AND AT LOW OR
-|- UNDETECTABLE LEVEL THEREAFTER.
-|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
-|- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
SIMILARITY).
-|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
-|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
-|- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; AF140673; AAD56356.1; -.
MEROPS; M12.225; -.
MGD; MGI:1346321; Adamts5.
InterPro; IPR001762; Disintegrin.
InterPro; IPR002870; Pep_M12B_propep.
InterPro; IPR001590; RepPolysin.
InterPro; IPR000884; TSP1.
InterPro; IPR000130; Zn MTpeptdse.
Pfam; PF00090; tsp 1; 2.
Pfam; PF01421; RepPolysin; 1.
Pfam; PF01562; Pep_M12B_propep; 1.
SMART; SM00209; TSP1; 2.
PROSITE; PS50215; ADAM_MEPRO; 1.
PROSITE; PS50092; TSP1; 1.
PROSITE; PS00142; ZINC_PROTEASE; 1.
PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 21
FT PROPEP 22 261
FT CHAIN 262 930
FT SITE 209 209
FT METAL 410 410
FT ACT_SITE 411 411
FT METAL 414 414
FT METAL 420 420
FT DOMAIN 485 566
FT DOMAIN 567 623
FT DOMAIN 624 731
FT DOMAIN 732 874
FT DOMAIN 875 930
FT DOMAIN 930 930
FT DOMAIN 930 930
FT DOMAIN 930 930
FT CARBOHYD 498 498
FT CARBOHYD 728 728
FT CARBOHYD 802 802
FT CARBOHYD 807 807
SQ SEQUENCE 930 AA; 101780 MW; 84DE84B26170D4DC CRC64;

Query Match 20.7%; Score 148.5; DB 1; Length 930;
Best Local Similarity 28.7%; Pred. No. 2.4e-07;
Matches 35; Conservative 22; Mismatches 64; Indels 1; Gaps 1;

```

Qy 15 ERYVLTNLNIGAEILRDPISLGAQPRVHLVKWVILTEPEGAPNITANLTSSLSVCGWSQT 74
Db 288 QHYLLTLASTANRLYSHASIENTHIRLAVKVVVLTDKDTSLEVSXNAATTLKNFCKWQH 347
Qy 75 INPEDDTPGHDVLVLYITRFDLELDPDGNRQVRGVTOLGACSPWTSCLITEDTGFDLGV 134
Db 348 HNQLGDDHEEHYDAAILFTREDL-CGHHSCDTLGMADVGTICSPERSCAVIEDDGLHA 406
Qy 135 TI 136
Db 407 TV 408

Search completed: March 20, 2003, 12:34:59
Job time : 21.5405 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:34:44 ; Search time 22.0541 Seconds
(without alignments)
329.684 Million cell updates/sec

Title: SEQID_15

Perfect score: 716
Sequence: 1 AVGPVDFQAHQEDTERYVLT.....SPTWSCLITDGFGLGVTI 136

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 221153

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW PUB.psp.*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW PUB.psp.*
 - 3: /cgn2_6/ptodata/2/pubpaa/US03_NEW PUB.psp.*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.psp.*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW PUB.psp.*
 - 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.psp.*
 - 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.psp.*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.psp.*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW PUB.psp.*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.psp.*
 - 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW PUB.psp.*
 - 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.psp.*
 - 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW PUB.psp.*
 - 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.psp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	716	100.0	136	10	US-09-833-328-15	Sequence 15, Appli
2	716	100.0	148	10	US-09-833-328-4	Sequence 4, Appli
3	716	100.0	150	10	US-09-833-328-6	Sequence 6, Appli
4	716	100.0	242	10	US-09-978-979-1	Sequence 1, Appli
5	716	100.0	1416	10	US-09-836-712-2	Sequence 2, Appli
6	713	99.6	203	10	US-09-781-080B-2	Sequence 2, Appli
7	713	99.6	1120	10	US-09-781-080B-11	Sequence 11, Appli
8	702	98.0	133	10	US-09-833-328-2	Sequence 2, Appli
9	212.5	29.7	997	10	US-09-918-171A-7	Sequence 7, Appli
10	206.5	28.8	1690	10	US-09-788-043C-5	Sequence 5, Appli
11	204.5	28.6	730	9	US-10-163-316-2	Sequence 16, Appli
12	201	28.1	2150	10	US-09-321-987B-2	Sequence 2, Appli
13	201	28.1	2165	10	US-09-800-729-155	Sequence 155, App
14	187.5	26.2	947	10	US-09-788-043C-7	Sequence 7, Appli
15	187.5	26.2	1629	10	US-09-972-467-2	Sequence 2, Appli
16	184	25.7	1882	10	US-09-918-171A-13	Sequence 13, Appli
17	180.5	25.2	839	9	US-09-912-788-2	Sequence 2, Appli
18	178.5	24.9	727	9	US-10-097-597-1	Sequence 1, Appli
19	178.5	24.9	727	9	US-10-097-580-1	Sequence 1, Appli

20	178.5	24.9	727	10	US-09-445-023A-1	Sequence 1, Appli
21	178.5	24.9	967	12	US-10-105-929-2	Sequence 2, Appli
22	177.5	24.8	727	9	US-10-097-597-12	Sequence 12, Appli
23	177.5	24.8	727	9	US-10-097-580-12	Sequence 12, Appli
24	177.5	24.8	727	10	US-09-445-023A-12	Sequence 12, Appli
25	177.5	24.8	950	10	US-09-321-987B-4	Sequence 4, Appli
26	177.5	24.8	968	9	US-10-163-316-7	Sequence 7, Appli
27	176	24.6	317	10	US-09-963-791-16	Sequence 16, Appli
28	176	24.6	356	10	US-09-963-791-20	Sequence 20, Appli
29	176	24.6	438	10	US-09-963-791-22	Sequence 22, Appli
30	176	24.6	468	10	US-09-963-791-6	Sequence 6, Appli
31	176	24.6	507	10	US-09-963-791-10	Sequence 10, Appli
32	176	24.6	589	10	US-09-963-791-12	Sequence 12, Appli
33	176	24.6	757	10	US-09-963-791-24	Sequence 24, Appli
34	176	24.6	859	10	US-09-918-171A-5	Sequence 5, Appli
35	176	24.6	908	10	US-09-963-791-2	Sequence 2, Appli
36	176	24.6	959	10	US-09-788-043C-1	Sequence 1, Appli
37	173.5	24.2	823	9	US-10-163-316-2	Sequence 2, Appli
38	173.5	24.2	823	9	US-09-965-631-4	Sequence 4, Appli
39	173.5	24.2	950	10	US-09-858-081-12	Sequence 12, Appli
40	172	24.0	814	10	US-09-788-043C-3	Sequence 3, Appli
41	172	24.0	947	10	US-09-858-081-9	Sequence 9, Appli
42	172	24.0	1044	10	US-09-858-081-2	Sequence 2, Appli
43	172	24.0	1068	10	US-09-858-081-2	Sequence 2, Appli
44	172	24.0	1133	10	US-09-858-068-2	Sequence 2, Appli
45	170	23.7	245	10	US-09-918-171A-11	Sequence 11, Appli

ALIGNMENTS

RESULT 1
US-09-833-328-15
; Sequence 15, Application US/09833328
; Patent No. US2002036713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Scheiflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Rndolf
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Gerlicsen, Helena E.
; TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protea
; TITLE OF INVENTION: Comprising a Polypeptide Chain with the Amino Acid Sequenc
; FILE REFERENCE: 247.00CIP
; CURRENT APPLICATION NUMBER: US/09/833.328
; PRIOR FILING DATE: 2001-04-12
; PRIOR FILING DATE: 09/721,254
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 15
; LENGTH: 136
; TYPE: PRT
; ORGANISM: human
US-09-833-328-15

Query Match 100.0%; Score 716; DB 10; Length 136;
Best Local Similarity 100.0%; Pred. No. 2e-73;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AVGPVDFQAHQEDTERYVLTNTNIGALLRDPGLGAQFRVHLVXKVILTEPEGAPNITAN 60
|||||
Db 1 AVGPVDFQAHQEDTERYVLTNTNIGALLRDPGLGAQFRVHLVXKVILTEPEGAPNITAN 60
QY 61 LTSSLLSVCGWSQTNPEDDTPGHADLVLYITRFDLELPDGNRVRGVTOLGGACSPTW 120
|||||
Db 61 LTSSLLSVCGWSQTNPEDDTPGHADLVLYITRFDLELPDGNRVRGVTOLGGACSPTW 120

```
QY 121 SCLITDTGFDLGVTI 136
Db 121 SCLITDTGFDLGVTI 136

RESULT 2
US-09-833-328-4
; Sequence 4, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwatz, Hans-Peter
; APPLICANT: Scheiflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Randolph
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Geritsen, Helena E.
; TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
; FILE REFERENCE: 247.00CIP
; CURRENT APPLICATION NUMBER: US/09/833,328
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 148
; TYPE: PRT
; ORGANISM: human
US-09-833-328-4

Query Match 100.0%; Score 716; DB 10; Length 148;
Best Local Similarity 100.0%; Pred. No. 2.3e-73;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVGPDVFOAQHEDTERYYVLTNLNIGAEALLRDPGLAQFRVHLVKMVLTEPEGAPNITAN 60
Db 13 AVGPDVFOAQHEDTERYYVLTNLNIGAEALLRDPGLAQFRVHLVKMVLTEPEGAPNITAN 72
QY 61 LTSSLLSVCGMSQTINPEDDTPGHADLVLYITRFDLELPGNRQVRGVTLGGACSPW 120
Db 73 LTSSLLSVCGMSQTINPEDDTPGHADLVLYITRFDLELPGNRQVRGVTLGGACSPW 132
QY 121 SCLITDTGFDLGVTI 136
Db 133 SCLITDTGFDLGVTI 148

RESULT 3
US-09-833-328-6
; Sequence 6, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwatz, Hans-Peter
; APPLICANT: Scheiflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Randolph
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Geritsen, Helena E.
; TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
; FILE REFERENCE: 247.00CIP
; CURRENT APPLICATION NUMBER: US/09/833,328
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254
```

```
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 150
; TYPE: PRT
; ORGANISM: human
US-09-833-328-6

Query Match 100.0%; Score 716; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 2.3e-73;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVGPDVFOAQHEDTERYYVLTNLNIGAEALLRDPGLAQFRVHLVKMVLTEPEGAPNITAN 60
Db 15 AVGPDVFOAQHEDTERYYVLTNLNIGAEALLRDPGLAQFRVHLVKMVLTEPEGAPNITAN 74
QY 61 LTSSLLSVCGMSQTINPEDDTPGHADLVLYITRFDLELPGNRQVRGVTLGGACSPW 120
Db 75 LTSSLLSVCGMSQTINPEDDTPGHADLVLYITRFDLELPGNRQVRGVTLGGACSPW 134
QY 121 SCLITDTGFDLGVTI 136
Db 135 SCLITDTGFDLGVTI 150

RESULT 4
US-09-978-979-1
; Sequence 1, Application US/09978979
; Patent No. US20020151702A1
; GENERAL INFORMATION:
; APPLICANT: Racie, Lisa, A.
; APPLICANT: Twine, Natalie, C.
; APPLICANT: Agostino, Michael, J.
; APPLICANT: Wolfman, Neil
; APPLICANT: Morris, Elisabeth
; TITLE OF INVENTION: Aggrecanase Molecules
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Home Products Corporation
; STREET: One Campus Drive
; CITY: Parsippany
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/978,979
; FILING DATE: 16-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/60/241,469
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Kapinos, Ellen J.
; REGISTRATION NUMBER: 32,345
; REFERENCE/DOCKET NUMBER: GI 5435p
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (973) 660-5000
; TELEFAX: (973) 683-4117
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 242 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-978-979-1
```

Query Match 100.0%; Score 716; DB 10; Length 242;
Best Local Similarity 100.0%; Pred. No. 4.4e-73;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVGPVFOAQHEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 60
DB 54 AVGPVFOAQHEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 113

QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 120
DB 114 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 173

QY 121 SCLITEDTGFDLGVTI 136
DB 174 SCLITEDTGFDLGVTI 189

RESULT 5
US-09-836-712-2
; Sequence 2, Application US/09836712
; Patent No. US20010049106A1
; GENERAL INFORMATION:
; APPLICANT: PRIZER INC.
; TITLE OF INVENTION: ADAMTS POLYPEPTIDES, NUCLEIC ACIDS ENCODING THEM, AND
; FILE REFERENCE: PC10851A
; CURRENT APPLICATION NUMBER: US/09/836,712
; CURRENT FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1416
; TYPE: PRT
; ORGANISM: Human
US-09-836-712-2

Query Match 100.0%; Score 716; DB 10; Length 1416;
Best Local Similarity 100.0%; Pred. No. 4.6e-72;
Matches 136; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVGPVFOAQHEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 60
DB 110 AVGPVFOAQHEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 169

QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 120
DB 170 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 229

QY 121 SCLITEDTGFDLGVTI 136
DB 230 SCLITEDTGFDLGVTI 245

RESULT 6
US-09-781-080B-2
; Sequence 2, Application US/09781080B
; Patent No. US20020142439A1
; GENERAL INFORMATION:
; APPLICANT: HOLLOWAY, JIM
; APPLICANT: SHEPPARD, PAUL
; TITLE OF INVENTION: Anti-Angiogenic Intestinal Peptides,
; FILE REFERENCE: zdint5
; CURRENT APPLICATION NUMBER: US/09/781,080B
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-781-080B-2

Query Match 99.6%; Score 713; DB 10; Length 203;
Best Local Similarity 99.3%; Pred. No. 7.5e-73;
Matches 135; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVGPVFOAQHEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 60
DB 14 AVGPVFOAQHEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 73

QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 120
DB 74 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 133

QY 121 SCLITEDTGFDLGVTI 136
DB 134 SCLITEDTGFDLGVTI 149

RESULT 7
US-09-781-080B-11
; Sequence 11, Application US/09781080B
; Patent No. US20020142439A1
; GENERAL INFORMATION:
; APPLICANT: HOLLOWAY, JIM
; APPLICANT: SHEPPARD, PAUL
; APPLICANT: YAMAMOTO, GAYLE
; TITLE OF INVENTION: Anti-Angiogenic Intestinal Peptides,
; FILE REFERENCE: zdint5
; CURRENT APPLICATION NUMBER: US/09/781,080B
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 1120
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1120)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-781-080B-11

Query Match 99.6%; Score 713; DB 10; Length 1120;
Best Local Similarity 99.3%; Pred. No. 7.4e-72;
Matches 135; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AVGPVFOAQHEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 60
DB 117 AVGPVFOAQHEDTERVLTNLNIGAEALLRDPGLGQAFRVHLVKMVLTEPEGAPNITAN 176

QY 61 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 120
DB 177 LTSSLLSVCGWSQTINPEDDTPGHADLVLYITRFDELPLDGNRQVRGVTLGGACSPW 236

QY 121 SCLITEDTGFDLGVTI 136
DB 237 SCLITEDTGFDLGVTI 252

RESULT 8
US-09-833-328-2
; Sequence 2, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Scheifflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Rndolf
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus

```
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Gexitsen, Helena E.
; TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
; TITLE OF INVENTION: Comprising a Polypeptide Chain with the Amino Acid Sequence AAGC
; FILE REFERENCE: 247.00CIP
; CURRENT APPLICATION NUMBER: US/09/833,328
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 133
; TYPE: PRT
; ORGANISM: human
US-09-833-328-2
```

Query Match 98.0%; Score 702; DB 10; Length 133;
Best Local Similarity 100.0%; Pred. No. 7.5e-72;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 4 PDVFOAHOEDTERVYLTNLNIGAEALLRDPISLGAQFRVHLVKMVLITPEGAPNITANLTS 63
Db 1 PDVFOAHOEDTERVYLTNLNIGAEALLRDPISLGAQFRVHLVKMVLITPEGAPNITANLTS 60
Qy 64 SLLSVCGMSQITNPEDDDTPGHADLVLYITRFDLELPGNRQVRGVTOLGACSPWTSCL 123
Db 61 SLLSVCGMSQITNPEDDDTPGHADLVLYITRFDLELPGNRQVRGVTOLGACSPWTSCL 120
Qy 124 ITEDTGFDLGVTI 136
Db 121 ITEDTGFDLGVTI 133
```

```
RESULT 9
US-09-918-171A-7
; Sequence 7, Application US/0991817A
; Patent No. US20020110894A1
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/04193
; CURRENT APPLICATION NUMBER: US/09/918,171A
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 09/369,364
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
US-09-918-171A-7
```

Query Match 29.7%; Score 212.5; DB 10; Length 997;
Best Local Similarity 37.5%; Pred. No. 2e-15;
Matches 48; Conservative 22; Mismatches 55; Indels 3; Gaps 2;

```
Qy 11 QEDTERVYLTNLNIGAEALLRDPISLGAQFRVHLVKMVLITPEGAPNITANLTSLLSVCG 70
Db 260 PQVESYVLTIMNVAGLFHDPISGNPIHITIVRLVLEDEEDLKITHADNTPKSFCK 319
Qy 71 WSQITNPEDDDTPGHADLVLYITRFDLELPGNR--QVRGVTOLGACSPWTSCLITEDT 128
Db 320 WQKSINMKGDAPHLPHDHTAILLRKDL-CAAMNRPCTGLSHVAGMCQPHRSCSINEDT 378
Qy 129 GFDLGVTI 136
Db 379 GLPLAFTV 386
```

```
RESULT 10
US-09-788-043C-5
; Sequence 5, Application US/09788043C
; Patent No. US20020107361A1
; GENERAL INFORMATION:
; APPLICANT: Heller, Renu
; APPLICANT: Zuo, Fengrong
; APPLICANT: Klonowski, Paul
; TITLE OF INVENTION: No. US20020107361A1e1 Metalloproteases Having
; TITLE OF INVENTION: Thrombospondin Domains and Nucleic Acid Compositions
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: ROCH-004
; CURRENT APPLICATION NUMBER: US/09/788,043C
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/184,152
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1690
; TYPE: PRT
; ORGANISM: human
US-09-788-043C-5

Query Match 28.8%; Score 206.5; DB 10; Length 1690;
Best Local Similarity 36.7%; Pred. No. 1.9e-14;
Matches 47; Conservative 22; Mismatches 56; Indels 3; Gaps 2;

Qy 11 QEDTERVYLTNLNIGAEALLRDPISLGAQFRVHLVKMVLITPEGAPNITANLTSLLSVCG 70
Db 264 PQVESYVLTIMNVAGLFHDPISGNPIHITIVRLVLEDEEDLKITHADNTPKSFCK 323
Qy 71 WSQITNPEDDDTPGHADLVLYITRFDLELPGNR--QVRGVTOLGACSPWTSCLITEDT 128
Db 324 WQKSINMKGDAPHLPHDHTAILLRKDL-CATMNRPCETGLSHVAGMCQPHRSCSINEDT 382
Qy 129 GFDLGVTI 136
Db 383 GLPLAFTV 390

RESULT 11
US-10-163-547-16
; Sequence 16, Application US/10163547
; Publication No. US20030037350A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Myer, Rachel
; APPLICANT: Rudolf-Owen, Laura A.
; TITLE OF INVENTION: No. US20030037350A1e1 nucleic acid sequences encoding a
; TITLE OF INVENTION: human ubiquitin protease, lipase, dynamin, short chain
; TITLE OF INVENTION: dehydrogenase, and ADAM-TS Metalloprotease and uses
; FILE REFERENCE: 35800/247836(5800-243)
; CURRENT APPLICATION NUMBER: US/10/163,547
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 09/407,356
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 09/704,918
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: 09/435,311
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 09/796,100
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/185,503
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: 09/781,598
; PRIOR FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/182,009
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 09/782,952
```


Patent No. US20020068319A1
GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: 32 Human secreted proteins
FILE REFERENCE: P2044P1
CURRENT APPLICATION NUMBER: US/09/800,729
CURRENT FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: PCT/US00/26013
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: 60/155,709
PRIOR FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 217
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 155
LENGTH: 2165
TYPE: PRT
ORGANISM: Homo sapiens
US-09-800-729-155

Query Match 28.1%; Score 201; DB 10; Length 2165;
Best Local Similarity 33.6%; Pred. No. 1.1e-13;
Matches 44; Conservative 17; Mismatches 70; Indels 0; Gaps 0;
QY 6 VFQAHQEDTERRYVLTNIGAEILLRDPGLGQAFRVHLVKMVLTEPEGAPNITANLTSSL 65
DB 292 MYEYHGRSLEDYVLTFLSTVASIYRHQSLRASINVVVVKLVLTENAGPRITQNAOQTL 351
QY 66 LSVCGWSQTNIPEDDTPGHADLVLYITRFDLELPDGNRQVRGVLTOLGGACSPWWSCLIT 125
DB 352 QDFCRWQYYNDPDDSSVQHHDVAILLTRKDCIRSGKCDTLGLAEELGTMDCMQKSCAI 411
QY 126 EDTGFDLGVTI 136
DB 412 EDNGLSAAFTI 422

RESULT 14
US-09-788-043C-7
Sequence 7, Application US/09788043C
Patent No. US20020107361A1
GENERAL INFORMATION:
APPLICANT: Heller, Renu
APPLICANT: Zuo, Fengrong
APPLICANT: Klonowski, Paul
TITLE OF INVENTION: No. US20020107361A1 Metalloproteases Having
TITLE OF INVENTION: Thrombospondin Domains and Nucleic Acid Compositions
TITLE OF INVENTION: Encoding the Same
FILE REFERENCE: ROCH-004
CURRENT APPLICATION NUMBER: US/09/788,043C
CURRENT FILING DATE: 2001-02-16
PRIOR APPLICATION NUMBER: 60/184,152
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 947
TYPE: PRT
ORGANISM: human
US-09-788-043C-7

Query Match 26.2%; Score 187.5; DB 10; Length 947;
Best Local Similarity 30.2%; Pred. No. 1.2e-12;
Matches 39; Conservative 28; Mismatches 55; Indels 7; Gaps 2;
QY 10 HQEDTERRYVLTNIGAEILLRDPGLGQAFRVHLVKMVLTEPEGAPNITANLTSSL 69
DB 309 HGENLQHYLTLMISIVASIKDPSIGNLVINIVLVIHNEQDGPISFNAGTTLKNFC 368
QY 70 GWSQTNIPEDDTPG--HADLVLYITRFDLELPDGNRQVRGVLTOLGGACSPWWSCLIT 127
DB 369 QWQHSKN-----SPGGIHHDTAVLLTRQDICRAHDKCDTLGLAEELGTICDPYRSCSISED 423
QY 128 TGFDLGVTI 136

Query Match 28.1%; Score 204.5; DB 9; Length 730;
Best Local Similarity 34.6%; Pred. No. 1e-14;
Matches 44; Conservative 26; Mismatches 54; Indels 3; Gaps 2;
QY 12 EDTERYVLTNIGAEILLRDPGLGQAFRVHLVKMVLTEPEGAPNITANLTSSLVCGW 71
DB 265 ENVESYIITIMVVTGLFHNFSIGNAIIHVVRLLLEEEEOGLKIVHHAETLSSFCCK 324
QY 72 SQTNIPEDDTPGHADLVLYITRFDLELPDGNR--QVRGVLTOLGGACSPWWSCLIT 129
DB 325 QKSINPKSDDLNPVHRGVAVLLTRKDI-CAGFNRPCTETLGLSHLGMCPHRSNINEDSG 383
QY 130 FDLGVTI 136
DB 384 LPLAFTI 390

PRIOR FILING DATE: 2001-02-14
PRIOR APPLICATION NUMBER: 60/182,408
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: 09/496,005
PRIOR FILING DATE: 2000-02-01
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 730
TYPE: PRT
ORGANISM: Homo sapiens
US-10-163-547-16

Query Match 28.1%; Score 201; DB 10; Length 2150;
Best Local Similarity 33.6%; Pred. No. 1.1e-13;
Matches 44; Conservative 17; Mismatches 70; Indels 0; Gaps 0;
QY 6 VFQAHQEDTERRYVLTNIGAEILLRDPGLGQAFRVHLVKMVLTEPEGAPNITANLTSSL 65
DB 292 MYEYHGRSLEDYVLTFLSTVASIYRHQSLRASINVVVVKLVLTENAGPRITQNAOQTL 351
QY 66 LSVCGWSQTNIPEDDTPGHADLVLYITRFDLELPDGNRQVRGVLTOLGGACSPWWSCLIT 125
DB 352 QDFCRWQYYNDPDDSSVQHHDVAILLTRKDCIRSGKCDTLGLAEELGTMDCMQKSCAI 411
QY 126 EDTGFDLGVTI 136
DB 412 EDNGLSAAFTI 422

RESULT 13
US-09-800-729-155
Sequence 155, Application US/09800729
Patent No. US20020102210A1
GENERAL INFORMATION:
APPLICANT: Kimble, Judith E
APPLICANT: Bleiloch, Robert H
TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
FILE REFERENCE: 960296.95386
CURRENT APPLICATION NUMBER: US/09/321,987B
CURRENT FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: 60/087,170
PRIOR FILING DATE: 1998-05-29
PRIOR APPLICATION NUMBER: 60/129,023
PRIOR FILING DATE: 1999-04-13
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 2150
TYPE: PRT
ORGANISM: Caenorhabditis elegans
US-09-321-987B-2

Query Match 28.1%; Score 201; DB 10; Length 2150;
Best Local Similarity 33.6%; Pred. No. 1.1e-13;
Matches 44; Conservative 17; Mismatches 70; Indels 0; Gaps 0;
QY 6 VFQAHQEDTERRYVLTNIGAEILLRDPGLGQAFRVHLVKMVLTEPEGAPNITANLTSSL 65
DB 292 MYEYHGRSLEDYVLTFLSTVASIYRHQSLRASINVVVVKLVLTENAGPRITQNAOQTL 351
QY 66 LSVCGWSQTNIPEDDTPGHADLVLYITRFDLELPDGNRQVRGVLTOLGGACSPWWSCLIT 125
DB 352 QDFCRWQYYNDPDDSSVQHHDVAILLTRKDCIRSGKCDTLGLAEELGTMDCMQKSCAI 411
QY 126 EDTGFDLGVTI 136
DB 412 EDNGLSAAFTI 422

Query Match 28.1%; Score 201; DB 10; Length 2150;
Best Local Similarity 33.6%; Pred. No. 1.1e-13;
Matches 44; Conservative 17; Mismatches 70; Indels 0; Gaps 0;
QY 6 VFQAHQEDTERRYVLTNIGAEILLRDPGLGQAFRVHLVKMVLTEPEGAPNITANLTSSL 65
DB 292 MYEYHGRSLEDYVLTFLSTVASIYRHQSLRASINVVVVKLVLTENAGPRITQNAOQTL 351
QY 66 LSVCGWSQTNIPEDDTPGHADLVLYITRFDLELPDGNRQVRGVLTOLGGACSPWWSCLIT 125
DB 352 QDFCRWQYYNDPDDSSVQHHDVAILLTRKDCIRSGKCDTLGLAEELGTMDCMQKSCAI 411
QY 126 EDTGFDLGVTI 136
DB 412 EDNGLSAAFTI 422

Query Match 28.1%; Score 201; DB 10; Length 2150;
Best Local Similarity 33.6%; Pred. No. 1.1e-13;
Matches 44; Conservative 17; Mismatches 70; Indels 0; Gaps 0;
QY 6 VFQAHQEDTERRYVLTNIGAEILLRDPGLGQAFRVHLVKMVLTEPEGAPNITANLTSSL 65
DB 292 MYEYHGRSLEDYVLTFLSTVASIYRHQSLRASINVVVVKLVLTENAGPRITQNAOQTL 351
QY 66 LSVCGWSQTNIPEDDTPGHADLVLYITRFDLELPDGNRQVRGVLTOLGGACSPWWSCLIT 125
DB 352 QDFCRWQYYNDPDDSSVQHHDVAILLTRKDCIRSGKCDTLGLAEELGTMDCMQKSCAI 411
QY 126 EDTGFDLGVTI 136
DB 412 EDNGLSAAFTI 422

Db 424 SGLSTAFTI 432

RESULT 15

US-09-972-467-2
; Sequence 2, Application US/09972467
; Patent No. US20020090373A1
; GENERAL INFORMATION:
; APPLICANT: PFIZER INC.
; TITLE OF INVENTION: ADAMTS POLYPEPTIDES, NUCLEIC ACIDS ENCODING THEM, AND
; FILE OF INVENTION: USES THEREOF
; FILE REFERENCE: PC10850A
; CURRENT APPLICATION NUMBER: US/09/972.467
; CURRENT FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1629
; TYPE: PRT
; ORGANISM: Human
US-09-972-467-2

Query Match 26.2%; Score 187.5; DB 10; Length 1629;
Best Local Similarity 30.2%; Pred. No. 2.5e-12;
Matches 39; Conservative 28; Mismatches 55; Indels 7; Gaps 2;

QY 10 HQEDTERVLTNLNIGAEELLRDPGLGAQFRVHLVKMVLTEPEGAPNITANLTSSLLSVC 69
Db 309 HGENLOHYILTMSIVASYKDFSGNLIIVIVNLIVHNEODGPSISFNAOTTLKNFC 368

QY 70 GWSQTINPEDDTPG--HADLVLYITRFDLEPDGNGRQVRGVTQLGCACSPWWSCLITED 127
Db 369 QWOHKN-----SPGGIHHDVALLTRQDICRAHDKCDTLGLAELGTICDPYRSCSISED 423

QY 128 TGFDLGVTI 136
Db 424 SGLSTAFTI 432

Search completed: March 20, 2003, 12:36:38
Job time : 26.0541 sec

GenCore version 5.1.4.p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:32:34 ; Search time 2.18919 Seconds
(without alignments)
526.959 Million cell updates/sec

Title: SEQID_1
Perfect score: 57
Sequence: 1 AAGGILHLELLV 12
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:
1: pir1:
2: pir2:
3: pir3:
4: pir4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	70.2	489	1 S36390	mitochondrial proc
2	38	66.7	316	2 T08807	mitochondrial proc
3	38	66.7	514	2 A31643	cell adhesion 80K
4	38	66.7	514	2 A44100	cell adhesion mole
5	38	66.7	526	2 S43561	YCR37C homolog K08
6	38	66.7	534	2 G88575	protein K08E5.2 li
7	38	66.7	1081	2 T31094	surface antigen B8
8	37	64.9	116	2 A72502	hypothetical prote
9	37	64.9	356	2 B90387	hypothetical prote
10	37	64.9	378	2 S12190	mobl protein - Thi
11	37	64.9	876	2 D70971	hypothetical prote
12	36	63.2	233	2 T35594	hypothetical prote
13	36	63.2	392	2 A44167	aminomethyltransf
14	36	63.2	480	2 A48043	ubiquinol-cytochro
15	36	63.2	505	2 C98300	hypothetical prote
16	36	63.2	505	2 AD2983	hypothetical prote
17	36	63.2	1096	2 A26879	pullulanase (EC 3.
18	36	63.2	1142	2 S62714	phytochrome lb - M
19	36	63.2	1178	2 S57535	probable membrane
20	36	63.2	1829	2 T14280	RW1 protein - mous
21	35	61.4	145	1 UOCA1	urotensin I precu
22	35	61.4	165	2 G90142	hypothetical prote
23	35	61.4	189	2 A32840	hypothetical 22K p
24	35	61.4	305	2 T15710	hypothetical prote
25	35	61.4	362	1 ZP80C1	ubiquinol-cytochro
26	35	61.4	385	2 AH2200	two-component hybr
27	35	61.4	458	2 H82200	sodium/alanine sym
28	35	61.4	691	2 S75863	translation elonga
29	35	61.4	692	2 AC2348	translation elonga

30	35	61.4	694	1 S04429	translation elonga
31	35	61.4	697	2 S04390	translation elonga
32	35	61.4	698	2 E75536	translation elonga
33	35	61.4	805	2 T48840	transducer protein
34	35	61.4	921	2 D86293	PH2-22 protein -
35	35	61.4	1038	2 S38801	pullulanase (EC 3.
36	35	61.4	1090	2 S11823	pullulanase (EC 3.
37	34	59.6	204	2 A71368	hypothetical prote
38	34	59.6	214	2 G83882	hypothetical prote
39	34	59.6	359	2 A2891	fix2 protein - Rhi
40	34	59.6	360	2 F96902	aspartate-semialde
41	34	59.6	382	2 T20655	hypothetical prote
42	34	59.6	399	2 AE3556	alanine racemase (
43	34	59.6	401	2 A75388	NADH dehydrogenase
44	34	59.6	403	2 D87597	hypothetical prote
45	34	59.6	424	2 T35924	hypothetical prote

ALIGNMENTS

RESULT 1

S36390
Mitochondrial processing peptidase (EC 3.4.24.64) beta chain precursor [validated]
N:Alternate names: beta-MPP; mitochondrial processing peptidase chain P52
C:Species: Rattus norvegicus (Norway rat)
C:Date: 09-Dec-1993 #sequence revision 14-Jul-1994 #text_change 31-Dec-2000
C:Accession: S36390; S36391; F01229
R:Pages, V.; Rosenberg, L.E.; Fenton, W.A.; Kalousek, F.
Proc. Natl. Acad. Sci. U.S.A. 90, 5355-5358, 1993
A:Title: The beta subunit of the mitochondrial processing peptidase from rat liver
A:Reference number: S36390; MUID:93281757; PMID:8506385
A:Accession: S36390
A:Molecule type: mRNA
A:Residues: 1-489 <PAC>
A:Cross-references: GB:L12965; NID:G294588; PIDN:AAA41633.1; PID:G294589
A:Accession: S36391
A:Molecule type: Protein
A:Residues: 46-67;92-104;116-125;409-423;432-441 <PA2>
C:Genetics:
A:Genome: nuclear
C:Superfamily: mitochondrial processing peptidase alpha chain
C:Keywords: heterodimer; hydrolase; metalloprotease; mitochondrial matrix; mitoc
F:1-45/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F:46-489/Product: mitochondrial processing peptidase beta chain #status experiment
F:101,105/Binding site: zinc (His) #status predicted
F:104/Active site: Glu #status predicted

Query Match 70.2%; Score 40; DB 1; Length 489;
Best Local Similarity 72.7%; Pred. No. 9.7; Indels 0; Gaps 0;
Matches 8; Conservative 1; Mismatches 2;

QY 1 AAGGILHLELL 11
||||: ||||
Db 243 AAGGVCHNELL 253

RESULT 2

T08807
Mitochondrial processing peptidase (EC 3.4.24.64) beta chain (similarity) - human
N:Alternate names: protein DKF2p586i1223.1
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 02-Feb-2001
C:Accession: T08807
R:Ansoorge, W.; Winkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, May 1999
A:Reference number: 216472
A:Accession: T08807
A:Molecule type: mRNA
A:Residues: 1-316 <ANS>
A:Cross-references: EMBL:AL050216
A:Experimental source: adult uterus; clone DKF2p586i1223
C:Genetics:

A;Note: DXFZ58611223.1

C;Superfamily: mitochondrial processing peptidase alpha chain

C;Keywords: hydrolase; metalloproteinase; mitochondrion; oxidative phosphorylation; oxid

Query Match 66.7%; Score 38; DB 2; Length 316;

Best Local Similarity 72.7%; Pred. No. 15; Mismatches 0; Gaps 0;

Matches 8; Conservative 1; Indels 2; Indels 0; Gaps 0;

QY 1 AAGGILHLELL 11

|||||

Db 70 AAGGVSHDELL 80

RESULT 3

A31643

cell adhesion 80K glycoprotein precursor - slime mold (Dictyostelium discoideum)

C;Alternate names: cell adhesion molecule csa; cell surface glycoprotein, 80K

C;Species: Dictyostelium discoideum

A;Variety: strain AX-2

C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 01-Dec-2000

C;Accession: S22066; A31643; A26310; A23951; A33463; S01473

R;Desbarats, L.; Lam, T.Y.; Wong, L.M.; Siu, C.H.

submitted to the EMBL Data Library, June 1992

A;Description: Identification of a unique camp-responsive element in the gene encoding

A;Reference number: S22066

A;Accession: S22066

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-514 <DES>

A;Cross-references: EMBL:X66483; NID:G7289; PIDN:CAA47110.1; PID:G7290

R;Kamboj, R.K.; Wong, L.M.; Lam, T.Y.; Siu, C.H.

J. Cell Biol. 107, 1835-1843, 1988

A;Title: Mapping of a cell-binding domain in the cell adhesion molecule gp80 of Dictyost

A;Reference number: A31643; MUID:89034443; PMID:3182938

A;Accession: A31643

A;Molecule type: mRNA

A;Residues: 1-514 <KAM>

A;Cross-references: GB:M36545; NID:G167803; PIDN:AAA33212.1; PID:G167804

R;Noegel, A.; Gerisch, G.; Stadler, J.; Westphal, M.

EMBO J. 5, 1473-1476, 1986

A;Title: Complete sequence and transcript regulation of a cell adhesion protein from agg

A;Reference number: A26310

A;Accession: A26310

A;Molecule type: mRNA

A;Residues: 1-215, 'G', 217-514 <NOE>

A;Cross-references: EMBL:X04004; NID:G7246; PIDN:CAA27634.1; PID:G671860

R;Wong, L.M.; Siu, C.H.

Proc. Natl. Acad. Sci. U.S.A. 83, 4248-4252, 1986

A;Title: Cloning of cDNA for the contact site A glycoprotein of Dictyostelium discoideum

A;Reference number: A23951

A;Accession: A23951

A;Molecule type: mRNA

A;Residues: 20-49 <WON>

A;Cross-references: GB:M13546; NID:G167699; PIDN:AAA33181.1; PID:G167700

R;Kamboj, R.K.; Garlepy, J.; Siu, C.H.

Cell 59, 615-625, 1989

A;Title: Identification of an octapeptide involved in homophilic interaction of the cell

A;Reference number: A33463; MUID:90058644; PMID:2582489

A;Accession: A33463

A;Molecule type: protein

A;Residues: 132-139 <KAZ>

R;Stadler, J.; Bordier, C.; Lottspeich, F.; Henschen, A.; Gerisch, G.

Hoppe-Seyler's Z. Physiol. Chem. 363, 771-776, 1982

A;Title: Improved purification and N-terminal amino acid sequence determination of the c

A;Reference number: S01473; MUID:83005404; PMID:7118072

A;Accession: S01473

A;Molecule type: protein

A;Residues: 20-37, 'X', 39-41, 'X', 43-46 <STA>

R;Stadler, J.; Keenan, T.W.; Bauer, G.; Gerisch, G.

EMBO J. 8, 371-377, 1989

A;Title: The contact site A glycoprotein of Dictyostelium discoideum carries a phospholi

A;Reference number: A56857; MUID:89251561; PMID:2721485

A;Contents: annotation; detection of glycosphingolipidinositol anchor

A;Query Match 66.7%; Score 38; DB 2; Length 526;

Best Local Similarity 80.0%; Pred. No. 26;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 GGILHLELLV 12

A;Note: no phosphoserine was detected but phosphoinositol and ceramide were
A;Note: the N-linked carbohydrates are sulfated type 1 oligosaccharides
C;Comment: This protein is involved in the formation of intercellular contacts upo

C;Genetics:

A;Gene: csa

C;Keywords: blocked carboxyl end; cell adhesion; glycoprotein; lipoprotein; membra

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-514/Product: cell adhesion 80K glycoprotein #status predicted <MAT>

F;132-139/Region: cell adhesion #status predicted

F;460-479/Region: 10-residue repeats

F;493-514/Domain: carboxyl-terminal propeptide #status predicted <CTP>

F;128,137,207,294,399/Binding site: carboxylate (Asn) (covalent) #status predicted

F;492/Modified site: GSI-anchor ethanolamine amidated carboxyl end (Ser) (in matur

Query Match 66.7%; Score 38; DB 2; Length 514;

Best Local Similarity 70.0%; Pred. No. 25;

Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 AGGILHLELL 11

|||||

Db 165 AGGLNLELI 174

RESULT 4

A44100

cell adhesion molecule gp80 - slime mold (Dictyostelium discoideum)

C;Species: Dictyostelium discoideum

C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994

C;Accession: A44100

R;Desbarats, L.; Lam, T.Y.; Wong, L.M.; Siu, C.H.

J. Biol. Chem. 267, 19655-19664, 1992

A;Title: Identification of a unique camp-response element in the gene encoding the

A;Reference number: A44100; MUID:92406924; PMID:1326559

A;Contents: AX2

A;Accession: A44100

A;Status: preliminary

A;Molecule type: nucleic acid

A;Residues: 1-514 <DES>

A;Note: sequence inconsistent with the nucleotide translation

A;Note: sequence extracted from NCBI backbone (NCBI:P113993)

Query Match 66.7%; Score 38; DB 2; Length 514;

Best Local Similarity 70.0%; Pred. No. 25;

Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 AGGILHLELL 11

|||||

Db 165 AGGLNLELI 174

RESULT 5

S43561

YCR37C homolog K08E5.2 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 07-Sep-1994 #sequence_revision 10-Nov-1995 #text_change 24-Nov-1999

C;Accession: S43561

R;Kerhaw, J.

submitted to the EMBL Data Library, March 1994

A;Reference number: S43561

A;Accession: S43561

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-526 <KER>

A;Cross-references: EMBL:Z30974

C;Genetics:

A;Introns: 33/3; 122/2; 169/2; 214/2; 253/2; 481/3

C;Superfamily: sodium/sulfate cotransporter

Query Match 66.7%; Score 38; DB 2; Length 526;

Best Local Similarity 80.0%; Pred. No. 26;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

Db      485 GGIISLELLV 494
||||:|||||
RESULT 6
G88575
A:Accession: A72502
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-116 <KAW>
A:Cross-references: DDBJ:AP000063; NID:G5105654; PIDN:BAA81001.1; PID:1044787; PI
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: G88575
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites Genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: G88575
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-534 <STO>
A:Cross-references: GB:chr_III; PIDN:CAA83225.1; PID:G3878357; GSPDB:GN00021; CESP:K08E5
C:Genetics:
A:Gene: K08E5.2
A:Map position: 3
C:Superfamily: sodium/sulfate cotransporter

Query Match      66.7%; Score 38; DB 2; Length 534;
Best Local Similarity 80.0%; Pred. No. 26;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 GGILHLELLV 12
||||:|||||
Db      493 GGIISLELLV 502

RESULT 7
T31094
surface antigen BspA - Bacteroides forsythus
C:Species: Bacteroides forsythus
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T31094
R:Sharma, A.; Sojar, H.T.; Glurich, I.; Honma, K.; Kuramitsu, H.K.; Genco, R.J.
Infect. Immun. 66, 5703-5710, 1998
A:Title: Cloning, expression, and sequencing of a cell surface antigen containing a leuc
A:Reference number: Z20977; MUID:99043895; PMID:9826345
A:Accession: T31094
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1081 <SHA>
A:Cross-references: EMBL:AF054892; NID:G3005672; PID:G3005673; PIDN:AAC82625.1
C:Genetics:
A:Gene: bspA

Query Match      66.7%; Score 38; DB 2; Length 1081;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAGGILHLEL 10
||||:|||||
Db      1023 AAGGALHLLT 1032

RESULT 8
A72502
hypothetical protein APE1991 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C:Accession: A72502
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Rep. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450; MUID:99310339; PMID:10382966

A:Accession: A72502
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-116 <KAW>
A:Cross-references: DDBJ:AP000063; NID:G5105654; PIDN:BAA81001.1; PID:1044787; PI
C:Genetics:
C:Superfamily: Aeropyrum pernix hypothetical protein APE1991

Query Match      64.9%; Score 37; DB 2; Length 116;
Best Local Similarity 77.8%; Pred. No. 7.5;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAGGILHLE 9
||||:|||||
Db      44 AAGGLOHLE 52

RESULT 9
B90387
hypothetical protein asd-2 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
C:Accession: B90387
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.;
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Red
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: B90387
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-356 <KUR>
A:Cross-references: GB:AE006641; NID:G13815477; PIDN:AAK42353.1; GSPDB:GN00155
C:Genetics:
A:Gene: asd-2
C:Superfamily: yeast aspartate-semialdehyde dehydrogenase

Query Match      64.9%; Score 37; DB 2; Length 356;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 AAGGILHLELLV 12
||||:|||||
Db      337 ACGGILATELLI 348

RESULT 10
S12190
mobL protein - Thiobacillus ferrooxidans plasmid pTF1
C:Species: Thiobacillus ferrooxidans
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S12190
R:Drolet, M.; Zanga, P.; Lau, P.C.K.
Mol. Microbiol. 4, 1381-1391, 1990
A:Title: The mobilization and origin of transfer regions of a Thiobacillus ferroox
A:Reference number: S12188; MUID:91125140; PMID:2280689
A:Accession: S12190
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-378 <DRO>
A:Cross-references: EMBL:X52699; NID:G48158; PIDN:CAA36927.1; PID:G48161
C:Genetics:
A:Gene: mobL
A:Genome: plasmid pTF1

Query Match      64.9%; Score 37; DB 2; Length 378;
Best Local Similarity 60.0%; Pred. No. 28;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      3 GGILHLELLV 12

```

```
Db 46 GCVLHSELIL 55
|||||
translational initiation region on expression.
A;Reference number: A44167; MUID:92406724; PMID:1526969
A;Accession: A44167
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-392 <OKA>
A;Cross-references: GB:D11162; NID:G222867; PIDN:BAA01937.1; PID:G222868
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIN:113296, NCBIP:113297)
A;Note: parts of this sequence, including the amino end of the mature protein, wa
R;Okamura-Ikeda, K.; Fujiwara, K.; Yamamoto, M.; Hiraga, K.; Motokawa, Y.
J. Biol. Chem. 266, 4917-4921, 1991
A;Title: Isolation and sequence determination of cDNA encoding T-protein of the gl
A;Reference number: A23707; MUID:91161577; PMID:2002038
A;Accession: B23707
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 179-392 <OK2>
A;Note: part of this sequence was confirmed by protein sequencing
C;Comment: A mutation of the T-protein in humans leads to nonketotic hypertycinem
C;Superfamily: aminomethyltransferase
C;Keywords: mitochondrion; transferase
F;1-16/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F;17-392/Product: aminomethyltransferase #status predicted <NAT>

Query Match 63.2%; Score 36; DB 2; Length 392;
Best Local Similarity 70.0%; Pred. No. 46;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGGILHLEL 10
|||||
Db 154 AAGGDVHLEV 163

RESULT 14
A48043
ubiquinol-cytochrome-c reductase (EC 1.10.2.2) core protein I - human
N;Alternate names: cytochrome bc1 core I protein
C;Species: Homo sapiens (man)
C;Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 03-Jun-2002
C;Accession: A48043; I52367
R;Hoffman, G.G.; Lee, S.; Christiano, A.M.; Chung-Honet, L.C.; Cheng, W.; Katchma
J. Biol. Chem. 268, 21113-21119, 1993
A;Title: Complete coding sequence, intron/exon organization, and chromosomal loca
A;Reference number: A48043; MUID:94012661; PMID:8407948
A;Accession: A48043
A;Molecule type: DNA
A;Residues: 1-480 <HOF>
A;Cross-references: GB:L16842; NID:G349472; PIDN:AAA20046.1; PID:G515634
R;Islam, M.M.; Tanaka, M.; Suzuki, H.; Torii, K.; Hattori, M.; Ozawa, T.
Biochem. Mol. Biol. Int. 32, 797-805, 1994
A;Title: A complete cDNA sequence for core I protein subunit of human ubiquinol-c
A;Reference number: I52367; MUID:94348430; PMID:8069229
A;Accession: I52367
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-300; N, 302-480 <ISL>
A;Cross-references: GB:D26485; NID:G468934; PIDN:BAA05495.1; PID:G468935
C;Genetics:
A;Gene: GDB:UO0RC1
A;Cross-references: GDB:141850; OMIM:191328
A;Map position: 3p21.3-3p21.2
C;Superfamily: mitochondrial processing peptidase alpha chain
C;Keywords: oxidoreductase

Query Match 63.2%; Score 36; DB 2; Length 480;
Best Local Similarity 63.6%; Pred. No. 57;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
|||||
Db 233 AAGGVEHQQLL 243

RESULT 13
A44167
aminomethyltransferase (EC 2.1.2.10) precursor - chicken
N;Alternate names: glycine cleavage system protein T
C;Species: Gallus gallus (chicken)
C;Date: 27-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000
C;Accession: A44167; B23707
R;Okamura-Ikeda, K.; Fujiwara, K.; Motokawa, Y.
J. Biol. Chem. 267, 18284-18290, 1992
A;Title: Molecular cloning of a cDNA encoding chicken T-protein of the glycine cleavage
```

```
Db 46 GCVLHSELIL 55
|||||
translational initiation region on expression.
A;Reference number: A44167; MUID:92406724; PMID:1526969
A;Accession: A44167
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-392 <OKA>
A;Cross-references: GB:D11162; NID:G222867; PIDN:BAA01937.1; PID:G222868
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIN:113296, NCBIP:113297)
A;Note: parts of this sequence, including the amino end of the mature protein, wa
R;Okamura-Ikeda, K.; Fujiwara, K.; Yamamoto, M.; Hiraga, K.; Motokawa, Y.
J. Biol. Chem. 266, 4917-4921, 1991
A;Title: Isolation and sequence determination of cDNA encoding T-protein of the gl
A;Reference number: A23707; MUID:91161577; PMID:2002038
A;Accession: B23707
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 179-392 <OK2>
A;Note: part of this sequence was confirmed by protein sequencing
C;Comment: A mutation of the T-protein in humans leads to nonketotic hypertycinem
C;Superfamily: aminomethyltransferase
C;Keywords: mitochondrion; transferase
F;1-16/Domain: transit peptide (mitochondrion) #status predicted <TNP>
F;17-392/Product: aminomethyltransferase #status predicted <NAT>

Query Match 63.2%; Score 36; DB 2; Length 392;
Best Local Similarity 70.0%; Pred. No. 46;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGGILHLEL 10
|||||
Db 154 AAGGDVHLEV 163

RESULT 14
A48043
ubiquinol-cytochrome-c reductase (EC 1.10.2.2) core protein I - human
N;Alternate names: cytochrome bc1 core I protein
C;Species: Homo sapiens (man)
C;Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 03-Jun-2002
C;Accession: A48043; I52367
R;Hoffman, G.G.; Lee, S.; Christiano, A.M.; Chung-Honet, L.C.; Cheng, W.; Katchma
J. Biol. Chem. 268, 21113-21119, 1993
A;Title: Complete coding sequence, intron/exon organization, and chromosomal loca
A;Reference number: A48043; MUID:94012661; PMID:8407948
A;Accession: A48043
A;Molecule type: DNA
A;Residues: 1-480 <HOF>
A;Cross-references: GB:L16842; NID:G349472; PIDN:AAA20046.1; PID:G515634
R;Islam, M.M.; Tanaka, M.; Suzuki, H.; Torii, K.; Hattori, M.; Ozawa, T.
Biochem. Mol. Biol. Int. 32, 797-805, 1994
A;Title: A complete cDNA sequence for core I protein subunit of human ubiquinol-c
A;Reference number: I52367; MUID:94348430; PMID:8069229
A;Accession: I52367
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-300; N, 302-480 <ISL>
A;Cross-references: GB:D26485; NID:G468934; PIDN:BAA05495.1; PID:G468935
C;Genetics:
A;Gene: GDB:UO0RC1
A;Cross-references: GDB:141850; OMIM:191328
A;Map position: 3p21.3-3p21.2
C;Superfamily: mitochondrial processing peptidase alpha chain
C;Keywords: oxidoreductase

Query Match 63.2%; Score 36; DB 2; Length 480;
Best Local Similarity 63.6%; Pred. No. 57;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
|||||
Db 233 AAGGVEHQQLL 243

RESULT 13
A44167
aminomethyltransferase (EC 2.1.2.10) precursor - chicken
N;Alternate names: glycine cleavage system protein T
C;Species: Gallus gallus (chicken)
C;Date: 27-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000
C;Accession: A44167; B23707
R;Okamura-Ikeda, K.; Fujiwara, K.; Motokawa, Y.
J. Biol. Chem. 267, 18284-18290, 1992
A;Title: Molecular cloning of a cDNA encoding chicken T-protein of the glycine cleavage
```

```
RESULT 15
C98300
hypothetical protein AGR_L_2721 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C;Accession: C98300
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: C98300
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-505 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89925.1; PID:gl5159878; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_2721
A;Map position: linear chromosome
C;Superfamily: phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent

Query Match          63.2%; Score 36; DB 2; Length 505;
Best Local Similarity 58.3%; Pred. No. 61;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAGGILHLELV 12
   |||||
DB 107 AAGGVAHLMCVV 118

Search completed: March 20, 2003, 12:35:30
Job time : 5.18919 secs
```

GenCore version 5.1.4.p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:32:11 ; Search time 1.45946 Seconds
(without alignment)
341.028 Million cell updates/sec

Title: SEQID_1

Perfect score: 57

Sequence: 1 AAGGILHLELLV 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	70.2	489	1 MPPB_MOUSE	Q9cxt8 mus musculus
2	40	70.2	489	1 MPPB_RAT	Q03346 rattus norv
3	38	66.7	489	1 MPPB_HUMAN	O75439 homo sapien
4	38	66.7	489	1 YN22_CAEEL	O21339 caenorhabdi
5	38	66.7	514	1 CSA_DICDI	P08796 dictyosteli
6	37	64.9	378	1 NOBL_THIFE	P20085 thlobacillu
7	36	63.2	392	1 GCST_CHICK	P28337 gallus gall
8	36	63.2	435	1 FIXC_RHISN	O53208 rhizobium s
9	36	63.2	461	1 NNPI_HUMAN	P56182 homo sapien
10	36	63.2	480	1 UCRI_HUMAN	P31930 homo sapien
11	36	63.2	480	1 UCRI_MOUSE	Q9c213 mus musculu
12	36	63.2	1096	1 PULA_KLEAE	P07811 klebsiella
13	36	63.2	1178	1 YNI7_YEAST	P48231 saccharomyc
14	36	63.2	1805	1 RW1_HUMAN	Q92545 homo sapien
15	36	63.2	1829	1 RW1_MOUSE	O70472 mus musculu
16	35	61.4	145	1 URI_CYPCA	P01146 cyprinus ca
17	35	61.4	189	1 YTRE_LEPBI	P20464 leptospira
18	35	61.4	480	1 UCRI_BOVIN	P31800 bos taurus
19	35	61.4	691	1 EFG2_SYNY3	P74228 synchocyst
20	35	61.4	694	1 EFG_SYNP6	P18667 synchococc
21	35	61.4	697	1 EFG_SPIPL	P13550 spirulina p
22	35	61.4	698	1 EFG_DEIRA	Q9rxk5 deinococcus
23	35	61.4	1090	1 PULA_KLEPN	P07266 klebsiella
24	34	59.6	355	1 Y816_DROME	Q9vat0 drosophila
25	34	59.6	359	1 FIXZ_RHILE	P07748 rhizobium 1
26	34	59.6	396	1 ALR_BRUME	Q8ydc3 brucella me
27	34	59.6	401	1 NUOD_DEIRA	Q9ru89 deinococcus
28	34	59.6	492	1 GVD2_HALN1	Q9hht2 halobacteri
29	34	59.6	492	1 GVD2_HALN2	P33958 halobacteri
30	34	59.6	545	1 ESTB_DROPS	P25726 drosophila
31	34	59.6	608	1 EX5A_ECOLI	P04993 escherichia
32	34	59.6	688	1 EFG_APPP	Q9zeu4 apple proli
33	34	59.6	700	1 EFG2_RALSO	Q9xrm7 ralstonia s

34	59.6	702	1	EFG_THICU	OS0565 thlobacillu
35	59.6	703	1	EFG1_RALSO	Q8xv10 ralstonia s
36	59.6	866	1	NASA_KLEPN	Q06457 klebsiella
37	59.6	1234	1	YNK5_CAEEL	P34578 caenorhabdi
38	57.9	118	1	ATP2_SYNP6	P08443 synchococc
39	57.9	239	1	TN14_HUMAN	O9qvh9 mus musculu
40	57.9	240	1	TN14_MOUSE	O43557 homo sapien
41	57.9	341	1	EFG_STTRA	P29341 streptomyce
42	57.9	363	1	BASS_ECOLI	P08442 escherichia
43	57.9	417	1	YAT1_SYNP6	P76198 escherichia
44	57.9	423	1	YDIN_ECOLI	Q9c124 pasteurella
45	57.9	434	1	OADB_PASMU	

ALIGNMENTS

RESULT 1
MPPB_MOUSE
ID MPPB_MOUSE STANDARD; PRT; 489 AA.
AC Q9CXT8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mitochondrial processing peptidase beta subunit, mitochondrial
DE precursor (EC 3.4.24.64) (Beta-MPP) (P-52).
GN MPPCB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Aehburner T., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.P.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guncich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtauki S.,
RA Hayashizaki Y.,
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: THE MITOCHONDRIAL PROCESSING PEPTIDASE CLEAVES
PRESEQUENCES FROM MITOCHONDRIAL PROTEIN PRECURSORS (BY
SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Release of N-terminal transit peptides from
precursor proteins imported into the mitochondrion, typically with
Arg in position P2.
CC -!- COFACTOR: REQUIRES DIVALENT CATIONS FOR ACTIVITY (BY SIMILARITY).
CC -!- SUBUNIT: HETERODIMER OF ALPHA AND BETA SUBUNITS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)
or send an email to license@ebi.ac.uk).


```

CC -----
DR EMBL; AK013995; BAB29105.1; -
DR MGD; MGI:120328; 3110004018Rik.
DR InterPro; IPR001431; Peptidase_M16.
DR Pfam; PF00675; Peptidase_M16; 1.
DR PROSITE; PS00143; INSULINASE; 1.
KW Hydrolase; Metalloprotease; Mitochondrion; Transit peptide.
FT TRANSIT 1 45 MITOCHONDRION (POTENTIAL).
FT CHAIN 46 489 MITOCHONDRIAL PROCESSING PEPTIDASE
FT METAL 101 101 ZINC (BY SIMILARITY).
FT ACT SITE 104 104 BY SIMILARITY.
FT METAL 105 105 ZINC (BY SIMILARITY).
FT METAL 105 105 ZINC (BY SIMILARITY).
SQ SEQUENCE 489 AA; 54614 MW; 1B594EE0B6FE34A4 CRC64;

Query Match 70.2%; Score 40; DB 1; Length 489;
Best Local Similarity 72.7%; Pred. No. 4.6;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
Db 243 AAGGVCHNELL 253

RESULT 2
MPPB_RAT STANDARD; PRT; 489 AA.
AC Q03346;
AT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mitochondrial processing peptidase beta subunit, mitochondrial
DE precursor (EC 3.4.24.64) (Beta-MPP) (p-52).
GN PMPCB OR MPPB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=91281757; PubMed=8506385;
RA Paces V., Rosenberg L.E., Fenton W.A., Kalousek F.;
RA "The beta subunit of the mitochondrial processing peptidase from rat
RT liver: cloning and sequencing of a cDNA and comparison with a
RT proposed family of metalloproteases."
RL Proc. Natl. Acad. Sci. U.S.A. 90:5355-5358 (1993).
RN [2]
RP SEQUENCE OF 3-489 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93135783; PubMed=8422255;
RA Kitada S., Niidome T., Nagano T., Ogishima T., Ito A.;
RA "Molecular cloning of the smaller subunit (p52) of rat liver
RT mitochondrial processing peptidase."
RL Biochem. Biophys. Res. Commun. 190:289-293 (1993).
RN [3]
RP REVISTONS.
RA Kitada S.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE MITOCHONDRIAL PROCESSING PROTEASE CLEAVES
CC PRESEQUENCES FROM MITOCHONDRIAL PROTEIN PRECURSORS.
CC -!- CATALYTIC ACTIVITY: Release of N-terminal transit peptides from
CC precursor proteins imported into the mitochondrion, typically with
CC Arg in position p2.
CC -!- COFACTOR: REQUIRES DIVALENT CATIONS FOR ACTIVITY.
CC -!- SUBUNIT: HETERODIMER OF ALPHA AND BETA SUBUNITS.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L12965; AAA11633.1; -
DR EMBL; D13907; BAA03007.1; -
DR PIR; PC1229; PC1229.
DR MEROPS; M16.003; -.
DR InterPro; IPR001431; Peptidase_M16.
DR Pfam; PF00675; Peptidase_M16; 1.
DR PROSITE; PS00143; INSULINASE; 1.
KW Hydrolase; Metalloprotease; Mitochondrion; Transit peptide.
FT TRANSIT 1 45 MITOCHONDRION (POTENTIAL).
FT CHAIN 46 489 MITOCHONDRIAL PROCESSING PEPTIDASE
FT METAL 101 101 ZINC (BY SIMILARITY).
FT ACT SITE 104 104 BY SIMILARITY.
FT METAL 105 105 ZINC (BY SIMILARITY).
FT CONFLICT 167 167 G -> R (IN REF. 2).
FT CONFLICT 312 312 R -> C (IN REF. 2).
SQ SEQUENCE 489 AA; 54318 MW; 98B56A8C62B9F852 CRC64;

Query Match 70.2%; Score 40; DB 1; Length 489;
Best Local Similarity 72.7%; Pred. No. 4.6;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
Db 243 AAGGVCHNELL 253

RESULT 3
MPPB_HUMAN STANDARD; PRT; 489 AA.
AC O75439;
AT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mitochondrial processing peptidase beta subunit, mitochondrial
DE precursor (EC 3.4.24.64) (Beta-MPP) (p-52).
GN PMPCB OR MPPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=98318631; PubMed=9653160;
RA Mao M., Fu G., Wu J.-S., Zhang Q.-H., Zhou J., Kan L.-H., Huang Q.-H.,
RA He K.-L., Gu B.-W., Han Z.-G., Shen Y., Gu J., Yu Y.-P., Xu S.-H.,
RA Wang Y.-X., Chen S.-J., Chen Z.;
RA "Identification of genes expressed in human CD34(+) hematopoietic
RT stem/progenitor cells by expressed sequence tags and efficient
RT full-length cDNA cloning."
RL Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180 (1998).
CC -!- FUNCTION: THE MITOCHONDRIAL PROCESSING PROTEASE CLEAVES
CC PRESEQUENCES FROM MITOCHONDRIAL PROTEIN PRECURSORS.
CC -!- CATALYTIC ACTIVITY: Release of N-terminal transit peptides from
CC precursor proteins imported into the mitochondrion, typically with
CC Arg in position p2.
CC -!- COFACTOR: REQUIRES DIVALENT CATIONS FOR ACTIVITY.
CC -!- SUBUNIT: HETERODIMER OF ALPHA AND BETA SUBUNITS.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```
CC -----
DR EMBL: AF054182; AAC39915.1; -.
DR MEROPS; M16.003; -.
DR Genew; HGNC:9119; PMPCB.
DR MIM; 603131; -.
DR InterPro; IPR001431; Peptidase_M16.
DR Pfam; PF00675; Peptidase_M16; 1.
DR PROSITE; PS00143; INSULINASE; 1.
KW Hydrolase; Metalloprotease; Mitochondrion; Transit peptide.
FT TRANSIT 1 45 MITOCHONDRIAL PROCESSING PEPTIDASE
FT CHAIN 46 489
FT METAL 101 101 ZINC (BY SIMILARITY).
FT ACT SITE 104 104 BY SIMILARITY.
FT METAL 105 105 ZINC (BY SIMILARITY).
SQ SEQUENCE 489 AA; 54167 MW; 440A25A3E4486D64 CRC64;

Query Match 66.7%; Score 38; DB 1; Length 489;
Best Local Similarity 72.7%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGGILHLELL 11
DB 243 AAGGVSHDELL 253

RESULT 4
YMW22_CAEEL STANDARD; PRT; 498 AA.
AC Q21339;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein K08E5.2 in chromosome III.
GN K08E5.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Kershaw J.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -1- SIMILARITY: BELONGS TO THE NADC/P/PHO87 FAMILY OF TRANSPORTERS.
CC NADC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z30974; CAA83225.2; -.
DR WormPep; K08E5.2; CE28598.
DR InterPro; IPR001898; Na_sulf_syp.
DR Pfam; PF00939; Na_sulf_syp; 1.
DR TIGRFAMs; TIGR00785; Gass; 1.
DR PROSITE; PS01271; NA_SULFATE; 1.
KW Hypothetical protein; Transmembrane; Transport.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 94 114
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 200 220 POTENTIAL.
FT TRANSMEM 261 281 POTENTIAL.
FT TRANSMEM 284 304 POTENTIAL.
```

```
FT TRANSMEM 332 352 POTENTIAL.
FT TRANSMEM 371 391 POTENTIAL.
FT TRANSMEM 428 448 POTENTIAL.
SQ SEQUENCE 498 AA; 54517 MW; FAEDB2FAA694BCB1 CRC64;

Query Match 66.7%; Score 38; DB 1; Length 498;
Best Local Similarity 80.0%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 GGILHLELLV 12
DB 457 GGILSLELLV 466

RESULT 5
CSA_DICDI STANDARD; PRT; 514 AA.
AC P08796; P19408;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Contact site A protein precursor (CSA) (Membrane-associated glycoprotein gp80) (Cell adhesion molecule gp80).
GN CSA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
[1]
RN SEQUENCE FROM N.A.
RA Noegel A., Gerlach G., Stadler J., Westphal M.;
RT "Complete sequence and transcript regulation of a cell adhesion protein from aggregating Dictyostelium cells.";
RL EMBO J. 5:1473-1476(1986).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=89105362; PubMed=3063296;
RA Su C.-H., Wong L.M., Lam T.Y., Kamboj R.K., Choi A., Cho A.;
RT "Molecular mechanisms of cell-cell interaction in Dictyostelium discoideum.";
RL Biochem. Cell Biol. 66:1089-1099(1988).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=AX2;
RX MEDLINE=92406924; PubMed=1326559;
RA Desbarats L., Lam T.Y., Wong L.M., Siu C.H.;
RT "Identification of a unique CAMP-response element in the gene encoding the cell adhesion molecule gp80 in Dictyostelium discoideum.";
RL J. Biol. Chem. 267:19655-19664(1992).
RN [4]
RN SEQUENCE OF 20-49.
RA Wong L.M., Siu C.-H.;
RT "Cloning of cDNA for the contact site A glycoprotein of Dictyostelium discoideum.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4248-4252(1986).
RN [5]
RN CELL-BINDING DOMAIN.
RX MEDLINE=9034443; PubMed=3182938;
RA Kamboj R.K., Wong L.M., Lam T.Y., Siu C.H.;
RT "Mapping of a cell-binding domain in the cell adhesion molecule gp80 of Dictyostelium discoideum.";
RL J. Cell Biol. 107:1835-1843(1988).
RN [6]
RN GPI-ANCHOR.
RX MEDLINE=89251561; PubMed=2721485;
RA Stadler J., Keenan T.W., Bauer G., Gerlach G.;
RT "The contact site A glycoprotein of Dictyostelium discoideum carries a phospholipid anchor of a novel type.";
RL EMBO J. 8:371-377(1989).
CC -1- FUNCTION: THIS CELL-SURFACE GLYCOPROTEIN MEDIATES CELL-CELL BINDING VIA HOMOPHILIC INTERACTION.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR THAT CONTAINS A PHOSPHOCERAMIDE MOIETY. SUCH ANCHOR MEDIATES A
```

CC FAST AND LONG PERSISTENCE CELL ADHESION OF THE PROTEIN.
CC -1- DEVELOPMENTAL STAGE: RESTRICTED TO THE AGGREGATION STAGE OF
CC DEVELOPMENT OF D.DISCOIDEUM.
CC -1- PTM: ACYLATED, PHOSPHORYLATED ON SERINE AND N-GLYCOSYLATED WITH
CC TWO TYPES OF OLIGOSACCHARIDE CHAINS.
CC -1- MISCELLANEOUS: THE EXPRESSION OF THIS STRINGENTLY REGULATED
CC PROTEIN DURING CELL DEVELOPMENT IS MEDIATED THROUGH CELL-SURFACE
CC CAMP RECEPTORS.
CC -1- SIMILARITY: THE C-TERMINAL REGION CONTAINS CLUSTERS OF PROLINE
CC REGULARLY ALTERNATING WITH A HYDROXYAMINO ACID AND SHOWS
CC SIMILARITY TO THE HINGE REGION OF IMMUNOGLOBULINS. THIS DOMAIN
CC MIGHT ACT AS A SPACER TO ELEVATE SITES ACTIVE IN CELL CONTACT
CC INTO THE EXTRACELLULAR SPACE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04004; CAA27634.1; -
DR EMBL; M36545; AAA33212.1; -
DR EMBL; X66483; CAA47110.1; -
DR PIR; A23951; A23951.
DR PIR; A26310; A26310.
DR PIR; A31643; A31643.
DR PIR; S22066; S22066.
DR PIR; A44100; A44100.
DR DictyDb; DD02005; csaa.
DR InterPro; IPR002909; IPT_TIG.
DR Pfam; PF01833; TIG; 3.
KW Cell adhesion; Glycoprotein; Membrane; Phosphorylation; Repeat;
KW GPI-anchor; Lipoprotein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 492
FT PROPEP 493 514
FT DOMAIN 20 453
FT DOMAIN 454 485
FT DOMAIN 462 479
FT REPEAT 462 469
FT REPEAT 472 479
FT CARBOHYD 128 128
FT CARBOHYD 137 137
FT CARBOHYD 207 207
FT CARBOHYD 294 294
FT CARBOHYD 399 399
FT LIPID 492 492
FT CONFLICT 216 216
SQ SEQUENCE 514 AA; 53696 MW; 96A1C7CF42FEC096 CRC64;
Query Match 66.7%; Score 38; DB 1; Length 514;
Best Local Similarity 70.0%; Pred. No. 12;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
OY 2 AGGILHLELL 11
DB 165 AGGLNLELI 174
RESULT 6
MOBL_THIFE
ID MOBL_THIFE STANDARD; PRT; 378 AA.
AC P20085;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE 01-FEB-1991 (Rel. 17, Last annotation update)
DE Mobilization protein mobL.
GN MOBL.
OS Thibacillus ferrooxidans.
OG Plasmid pTF1.
OC Bacteria; Proteobacteria; gamma subdivision; Acidithiobacillus.

OX NCBI_TaxID=920;
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 33020;
RX MEDLINE=91125140; PubMed=2280689;
RA Drole M., Zanga P., Lau P.C.K.;
RT "The mobilization and origin of transfer regions of a Thibacillus
RT ferrooxidans plasmid: relatedness to plasmids RSF1010 and pSC101.";
RL Microbiol. 4:1381-1391(1990).
CC -1- FUNCTION: THIS PROTEIN IS ESSENTIAL TO PROMOTE THE SPECIFIC
CC TRANSFER OF THE PLASMID IN THE PRESENCE OF CONJUGATIVE PLASMIDS.
CC -1- SIMILARITY: TO MOBILIZATION PROTEIN A OF THE E.COLI PLASMID INCO
CC RSF1010 AND TO THE 45 kDa RELAXATION PROTEIN OF THE S.TYPHIMURIUM
CC PLASMID PSC101.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52699; CAA36927.1; -
DR PIR; S12190; S12190.
DR InterPro; IPR005053; MobA_Mobl.
DR Pfam; PF03389; MobA_Mobl; 1.
KW Mobility protein; Plasmid; Trans-acting factor; Conjugation.
SQ SEQUENCE 378 AA; 42636 MW; 4AAE91530BC0C194 CRC64;
Query Match 64.9%; Score 37; DB 1; Length 378;
Best Local Similarity 60.0%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
OY 3 GGILHLELLV 12
DB 46 GGVLHSELIL 55
RESULT 7
GCST_CHICK
ID GCST_CHICK STANDARD; PRT; 392 AA.
AC P2837;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aminomethyltransferase, mitochondrial precursor (EC 2.1.2.10) (Glycine
DE cleavage system T protein) (GCVT).
GN AMT.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92406724; PubMed=1526969;
RA Okamura-Ikeda K., Fujiwara K., Motokawa Y.;
RT "Molecular cloning of a cDNA encoding chicken T-protein of the
RT glycine cleavage system and expression of the functional protein in
RT Escherichia coli. Effect of mRNA secondary structure in the
RT translational initiation region on expression.";
RL J. Biol. Chem. 267:18284-18290(1992).
RN [2]
RP SEQUENCE OF 179-392 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91161577; PubMed=2002038;
RA Okamura-Ikeda K., Fujiwara K., Yamamoto M., Hiraga K., Motokawa Y.;
RT "Isolation and sequence determination of cDNA encoding T-protein of
RT the glycine cleavage system.";
RL J. Biol. Chem. 266:4917-4921(1991).
CC -1- FUNCTION: The glycine cleavage system catalyzes the degradation of
CC glycine.


```

RX MEDLINE=99276573; PubMed=10341208;
RA Savino T.M., Batos R., Jansen E., Hernandez-Verdun D.;
RT "The nuclear antigen Nop52, the human homologue of the yeast
RT ribosomal RNA processing RRP1, is recruited at late stages of
RT nucleogenesis.";
RL J. Cell Sci. 112:1889-1900(1999).
CC -|- FUNCTION: PLAYS A CRITICAL ROLE IN THE GENERATION OF 28S RNA.
CC -|- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -|- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED IN FETAL AND ADULT
CC TISSUES.
CC -|- SIMILARITY: BELONGS TO THE NNP-1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U79775; AAC51625.1; -
DR EMBL; AP001752; BAA95542.1; -
DR EMBL; BC000380; AAH00380.1; -
DR EMBL; AL137757; CAB70909.1; -
KW Nuclear protein.
FT CONFLICT 326 326 K -> R (IN REF. 4).
SQ SEQUENCE 461 AA; 52839 MW; C6F204EF4161AEAE CRC64;

Query Match 63.2%; Score 36; DB 1; Length 461;
Best Local Similarity 72.7%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAGGILHLELL 11
Db 43 AAGGFTDELL 53

RESULT 10
UCR1 HUMAN STANDARD; PRT; 480 AA.
AC P31930;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ubiquinol-cytochrome C reductase complex core protein I, mitochondrial
DE precursor (EC 1.10.2.2).
GN UQCRC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=94012661; PubMed=8407948;
RA Hoffman G.G., Lee S., Christiano A.M., Chung-Honet L.C., Cheng W.,
RA Katchman S., Uitto J., Greenspan D.S.;
RT "Complete coding sequence, intron/exon organization, and chromosomal
RT location of the gene for the core I protein of human ubiquinol-
RT cytochrome c reductase.";
RT J. Biol. Chem. 268:21113-21119(1993).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT ASN-301.
RC TISSUE=Fibroblast;
RX MEDLINE=94348430; PubMed=8069229;
RA Islam M.M., Tanaka M., Suzuki H., Torii K., Hattori N., Ozawa T.;
RT "A complete cDNA sequence for core I protein subunit of human
RT ubiquinol-cytochrome c reductase.";
RL Biochem. Mol. Biol. Int. 32:797-805(1994).
RN [3]
RP ERRATUM.
RX MEDLINE=95038519; PubMed=7951059;
RA Islam M.M., Tanaka M., Suzuki H., Torii K., Hattori N., Ozawa T.;

```

```

RL Biochem. Mol. Biol. Int. 33:410-410(1994).
RN [4]
RP ERRATUM.
RX MEDLINE=95072552; PubMed=7981668;
RA Islam M.M., Tanaka M., Suzuki H., Torii K., Hattori N., Ozawa T.;
RL Biochem. Mol. Biol. Int. 33:815-815(1994).
RN [5]
RP SEQUENCE OF 35-45.
RC TISSUE=Liver;
RX MEDLINE=94147969; PubMed=8313870;
RA Hughes G.J., Frutiger S., Paquet N., Pasquali C., Sanchez J.-C.,
RA Tissot J.-D., Bairoch A., Appel R.D., Hochstrasser D.F.;
RT "Human liver protein map: update 1993.";
RL Electrophoresis 14:1216-1222(1993).
RN [6]
RP VARIANT ASN-301.
RX PubMed=10453733;
RA Valnot I., Kaasis J., Chretien D., de Lonlay P., Parfait B.,
RA Munnich A., Kachaner J., Rustin P., Roetig A.;
RT "A mitochondrial cytochrome b mutation but no mutations of nuclearly
RT encoded subunits in ubiquinol cytochrome c reductase (complex III)
RT deficiency.";
RL Hum. Genet. 104:460-466(1999).
CC -|- FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C
CC REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH
CC IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN. THIS PROTEIN MAY
CC BE INVOLVED IN THE FORMATION OF THE COMPLEX BETWEEN CYTOCHROMES C AND C1.
CC -|- CATALYTIC ACTIVITY: OH(2) + 2 ferricytochrome c = O + 2
CC ferrocycytochrome c.
CC -|- SUBUNIT: B1 COMPLEX CONTAINS 11 SUBUNITS; 3 RESPIRATORY SUBUNITS
CC (CYTOCHROME B, CYTOCHROME C1, RIESKE PROTEIN), 2 CORE PROTEINS AND
CC 6 LOW-MOLECULAR WEIGHT PROTEINS.
CC -|- SUBCELLULAR LOCATION: Mitochondrial inner membrane.
CC -|- CAUTION: DOES NOT SEEM TO HAVE A PROTEASE ACTIVITY AS IT LACK THE
CC ZINC-BINDING SITE.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L16842; AAA20046.1; -
DR EMBL; D26485; BAA05495.1; -
DR HSSP; P31800; 1BE3
DR MEROPS; M16_973; -
DR SWISS-2DPAGE; P31930; HUMAN.
DR PHC1-2DPAGE; P31930; -
DR HSC-2DPAGE; P31930; HUMAN.
DR Sienna-2DPAGE; P31930; -
DR Genew; HGNC:12585; UQCRC1.
DR MIM; 191328; -
DR InterPro; IPR001431; Peptidase M16.
DR Pfam; PF00675; Peptidase M16; 1.
DR PROSITE; PS00143; INSULINASE; FALSE NEG.
DR Mitochondrion; Inner membrane; Electron transport; Respiratory chain;
KW Oxidoreductase; Transit peptide; Polymorphism.
FT TRANSIT 1 34 MITOCHONDRION
FT CHAIN 35 480 UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX
FT VARIANT 301 301 CORE PROTEIN I.
FT S -> N.
FT /FTID=VAR 013629.
SQ SEQUENCE 480 AA; 52618 MW; AA6B152E0181679 CRC64;

Query Match 63.2%; Score 36; DB 1; Length 480;
Best Local Similarity 63.6%; Pred. No. 27;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGILHLELL 11
|||:|:|

```

```

Db 233 AAGGVEHQQLL 243
RESULT 11
UCR1_MOUSE
ID UCR1_MOUSE STANDARD; PRT; 480 AA.
AC Q9CZ13; Q9CWL6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ubiquinol-cytochrome C reductase complex core protein I, mitochondrial
DE precursor (EC 1.10.2.2).
GN UQCRC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schraml L.M., Staali F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barch G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayaehizaki Y.;
RT Functional annotation of a full-length mouse cDNA collection.;
RL Nature 409:685-690(2001).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C
CC REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH
CC IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN. THIS PROTEIN MAY
CC MEDIATE FORMATION OF THE COMPLEX BETWEEN CYTOCHROMES C AND C1 (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: QH(2) + 2 ferricytochrome c = Q + 2
CC ferrocycytochrome c.
CC -1- SUBUNIT: BCI COMPLEX CONTAINS 11 SUBUNITS, 3 RESPIRATORY SUBUNITS
CC (CYTOCHROME B, CYTOCHROME C1, RIESKE PROTEIN), 2 CORE PROTEINS AND
CC 6 LOW-MOLECULAR WEIGHT PROTEINS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane (By
CC similarity).
CC -1- CAUTION: DOES NOT SEEM TO HAVE A PROTEASE ACTIVITY AS IT LACK THE
CC ZINC-BINDING SITE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AK013128; BAB28666.1; -;
CC DR EMBL; AK010553; BAB27022.1; -;
CC DR HSP; P31800; 1BE3.
CC DR MGD; MGI-107876; Ugcrc1.
CC DR InterPro; IPR001433; Peptidase M16.
CC DR Pfam; PF006675; Peptidase M16; 1.
CC DR PROSITE; PS00143; INSULINASE; FALSE NEG.
CC Mitochondrion; Inner membrane; Electron transport; Respiratory chain;
KW Oxidoreductase; Transit peptide.
FT TRANSIT 1 34 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 35 480 UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX
FT CORE PROTEIN I.
FT CONFLICT 223 223 N -> H (IN REF. 1; BAB27022).
FT CONFLICT 318 318 C -> Y (IN REF. 1; BAB27022).
FT SEQUENCE 480 AA; 52768 MW; 9C6D480DC9D5E429 CRC64;
Query Match 63.2%; Score 36; DB 1; Length 480;
Best Local Similarity 63.6%; Pred. No. 27;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAGGILHLELL 11
|||||:|:|
Db 233 AAGGVEHQQLL 243
RESULT 12
PULA_KLEAE
ID PULA_KLEAE STANDARD; PRT; 1096 AA.
AC P07811;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pullulanase precursor (EC 3.2.1.41) (Alpha-dextrin endo-1,6-alpha-
DE glucosidase) (Pullulan 6-glucanohydrolase).
GN PULA.
OS Klebsiella aerogenes.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=28451;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W70;
RX MEDLINE=87194626; PubMed=3155373;
RA Katsuragi N., Takizawa N., Murooka Y.;
RT Entire nucleotide sequence of the pullulanase gene of Klebsiella
RT aerogenes W70.;
RL J. Bacteriol. 169:2301-2306(1987).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic
CC linkages in pullulan and in amylopectin and glycogen, and the
CC alpha- and beta-limit dextrins of amylopectin and glycogen.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Probable).
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M16187; AAA25124.1; ALT_SEQ.
CC PIR; A26879; A26879.
CC InterPro; IPR000461; Alpha_amlase.
CC InterPro; IPR004193; Isoamylase_N.
CC InterPro; IPR005323; PUD.
CC Pfam; PF00128; alpha-amyase; 1.
CC Pfam; PF02922; isoamylase_N; 1.
CC Pfam; PF03714; PUD; 1.
CC PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
CC Hydroxylase; Glycosidase; Membrane; Lipoprotein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 1096 PULLULANASE.
FT LIPID 20 20 N-ACYL DIGLYCERIDE.
FT ACT_SITE 694 694 BY SIMILARITY.
FT ACT_SITE 723 723 BY SIMILARITY.
FT ACT_SITE 851 851 BY SIMILARITY.
FT SEQUENCE 1096 AA; 119335 MW; FE7D9167CDACPD79 CRC64;

```

Query Match 63.2%; Score 36; DB 1; Length 1096;
 Best Local Similarity 54.5%; Pred. No. 63;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGTLHLELL 11
 DB 484 SASGVTHLELL 494

RESULT 13
 YN17 YEAST
 ID YN17 YEAST STANDARD; PRT; 1178 AA.

AC P48231;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical 132.5 kDa protein in TOP2-MKT1 intergenic region.
 GN YNL087W OR N2250.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / FY1679;
 RX MEDLINE=96310628; PubMed=8740422;
 RA Soler-Mira A., Saiz J.E., Ballesta J.P.G., Remacha M.;
 RT "The sequence of a 17,933 bp segment of Saccharomyces cerevisiae
 chromosome XIV contains the RHO2, TOP2, MKT1 and END3 genes and five
 new open reading frames.";
 RL Yeast 12:485-491(1996).
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -|- SIMILARITY: CONTAINS 3 C2 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X89016; CAA61423.1; -
 DR EMBL; Z71363; CAA95963.1; -
 DR HSP; P21707; 1BYN.
 DR SGD; S0005031; YNL087W.
 DR InterPro; IPR000008; C2.
 DR Pfam; PF00168; C2; 3.
 DR SMART; SM00239; C2; 4.
 DR PROSITE; PS50004; C2_DOMAIN_2; 3.
 KW Hypothetical protein; Transmembrane; Repeat.
 FT DOMAIN 380 463
 FT C2 DOMAIN 1.
 FT C2 DOMAIN 2.
 FT C2 DOMAIN 3.
 FT TRANSMEM 103 123
 FT POTENTIAL.
 FT TRANSMEM 272 292
 FT POTENTIAL.
 FT TRANSMEM 324 344
 FT POTENTIAL.
 FT TRANSMEM 766 786
 FT POTENTIAL.
 FT TRANSMEM 855 875
 FT POTENTIAL.
 FT TRANSMEM 1134 1154
 FT POTENTIAL.
 SQ SEQUENCE 1178 AA; 132509 MW; 6597DE191C739FA1 CRC64;

Query Match 63.2%; Score 36; DB 1; Length 1178;
 Best Local Similarity 85.7%; Pred. No. 68;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 GGTLHLE 9
 DB 1096 GGVTHLE 1102

RESULT 14

RW1_HUMAN
 ID RW1_HUMAN STANDARD; PRT; 1805 AA.
 AC Q92545;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE RMI protein (Fragment).
 GN RMI OR KIAA0257.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=97191544; PubMed=9039502;
 RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawabayashi Y.,
 RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. VI.
 RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
 RT analysis of cDNA clones from cell line KG-1 and brain.";
 RL DNA Res. 3:321-329(1996).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; D87446; BAA13387.1; -
 FT NON_TER 1
 SQ SEQUENCE 1805 AA; 197588 MW; 955589214413AB84 CRC64;

Query Match 63.2%; Score 36; DB 1; Length 1805;
 Best Local Similarity 70.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAGGTLHLEL 10
 DB 169 SSGDLHLEL 178

RESULT 15
 RW1_MOUSE
 ID RW1_MOUSE STANDARD; PRT; 1829 AA.
 AC O70472;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE RMI protein.
 GN RMI.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BAUB/c; TISSUE=Brain;
 RA Wilkinson R., Fitter S., Tacharke D., Simmons A.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 CC -|- FUNCTION: MAY PLAY A ROLE IN THE IMMUNE RESPONSE TO VIRAL
 CC INFECTION.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF060565; AAC15232.1; -.
DR MGD; MGI:1927110; Rwl.
SQ SEQUENCE 1829 AA; 200508 MW; DOB3E209257AFAB9 CRC64;
Query Match 63.2%; Score 36; DB 1; Length 1829;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAGGILHLEL 10
:|||||
Db 197 SSGDLHLEL 206

Search completed: March 20, 2003, 12:34:54
Job time : 4.45946 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2003, 12:33:39 ; Search time 1.94595 Seconds
(without alignments)
181.441 Million cell updates/sec

Title: SEQID_1

Perfect score: 57
Sequence: 1 AAGGILHLELLV 12

Scoring table: BLOSUM62.
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	70.2	489	2	US-08-895-521-3
2	40	70.2	489	3	US-09-235-218-3
3	38	66.7	489	2	US-08-895-521-1
4	38	66.7	489	3	US-09-235-218-1
5	36	63.2	480	2	US-08-895-521-4
6	36	63.2	480	3	US-09-235-218-4
7	36	63.2	611	3	US-08-904-871-4
8	36	63.2	1096	4	US-09-346-237-6
9	36	63.2	1142	3	US-08-904-871-12
10	35	61.4	1090	4	US-09-346-237-5
11	33	57.9	239	4	US-08-913-014A-2
12	33	57.9	239	4	US-08-913-014A-3
13	33	57.9	240	4	US-08-913-014A-1
14	33	57.9	240	4	US-09-072-993C-4
15	33	57.9	242	4	US-08-913-014A-25
16	33	57.9	743	4	US-09-077-354B-2
17	32	56.1	170	3	US-08-110-663-2
18	32	56.1	170	4	US-08-432-335-2
19	32	56.1	170	4	US-09-614-022-2
20	32	56.1	237	3	US-09-248-335-48
21	32	56.1	343	4	US-09-413-814-99
22	32	56.1	429	4	US-08-965-762-29
23	32	56.1	429	4	US-09-911-927-29
24	32	56.1	429	4	US-09-911-882-29
25	32	56.1	656	4	US-09-605-785-379
26	32	56.1	656	4	US-09-439-313-379
27	32	56.1	656	4	US-09-352-616A-379

Sequence 380, App
Sequence 380, App
Sequence 380, App
Sequence 378, App
Sequence 378, App
Sequence 378, App
Sequence 378, App
Sequence 27, Appli
Sequence 2, Appli
Sequence 118, App
Sequence 118, App
Sequence 118, App
Sequence 4740, Ap
Sequence 6, Appli
Sequence 9, Appli
Sequence 6, Appli
Sequence 9, Appli
Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-895-521-3
; Sequence 3, Application US/08895521
; Patent No. 5869311
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
; TITLE OF INVENTION: SUBUNIT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,521
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0343 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 489 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 294589
; US-08-895-521-3

Query Match 70.2%; Score 40; DB 2; Length 489;

Best Local Similarity 72.7%; Pred. No. 16;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
||||:|
Db 243 AAGGVCHNELL 253

RESULT 2

US-09-235-218-3
; Sequence 3, Application US/09235218
; Patent No. 6001629
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
; TITLE OF INVENTION: SUBUNIT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,218
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/895,521
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0343 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 294589
US-09-235-218-3

Query Match 70.2%; Score 40; DB 3; Length 489;
Best Local Similarity 72.7%; Pred. No. 16;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
||||:|
Db 243 AAGGVCHNELL 253

RESULT 3

US-08-895-521-1
; Sequence 1, Application US/08895521
; Patent No. 5869311
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi

APPLICANT: Corley, Neil C.
TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
TITLE OF INVENTION: SUBUNIT
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895,521
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0343 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAINOT14
CLONE: 1593490
US-08-895-521-1

Query Match 66.7%; Score 38; DB 2; Length 489;
Best Local Similarity 72.7%; Pred. No. 37;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAGGILHLELL 11
||||:|
Db 243 AAGGVSHDELL 253

RESULT 4

US-09-235-218-1
; Sequence 1, Application US/09235218
; Patent No. 6001629
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
; TITLE OF INVENTION: SUBUNIT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

```
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/235,218
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/895,521
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0343 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 489 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAINOT14
CLONE: 1593490
US-09-235-218-1

Query Match      66.7%; Score 38; DB 3; Length 489;
Best Local Similarity 72.7%; Pred. No. 37;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY      1 AAGGILHLELL 11
      |||||:||||
Db      243 AAGGVSHDELL 253

RESULT 5
US-08-895-521-4
Sequence 4, Application US/08895521
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
TITLE OF INVENTION: SUBUNIT
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895,521
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0343 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
```

```
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1082896
US-08-895-521-4

Query Match      63.2%; Score 36; DB 2; Length 480;
Best Local Similarity 63.6%; Pred. No. 82;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY      1 AAGGILHLELL 11
      |||||:||||
Db      233 AAGGVHQLL 243

RESULT 6
US-09-235-218-4
Sequence 4, Application US/09235218
Patent No. 6001629
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
TITLE OF INVENTION: MITOCHONDRIAL PROCESSING PEPTIDASE
TITLE OF INVENTION: SUBUNIT
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/235,218
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/895,521
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0343 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1082896
US-09-235-218-4

Query Match      63.2%; Score 36; DB 3; Length 480;
Best Local Similarity 63.6%; Pred. No. 82;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

OY 1 AAGGILHELL 11
|||:|:|
Db 233 AAGGVEHOLL 243

RESULT 7
US-08-904-871-4
; Sequence 4, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C
; APPLICANT: Murphy, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134USO UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; CURRENT FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; EARLIER FILING DATE: 1996-08-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:: Sequence
US-08-904-871-4

Query Match 63.2%; Score 36; DB 3; Length 611;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILH 7
|||
Db 355 AAGGILH 361

RESULT 8
US-09-346-237-6
; Sequence 6, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Blegard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1096
; TYPE: PRT
; ORGANISM: Klebsiella aerogenes
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(1096)
; OTHER INFORMATION: Pullulanase
US-09-346-237-6

Query Match 63.2%; Score 36; DB 4; Length 1096;
Best Local Similarity 54.5%; Pred. No. 2e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGGILHELL 11
|||:|:|
Db 484 SASGVTHIELL 494

RESULT 9
US-08-904-871-12
; Sequence 12, Application US/08904871
; Patent No. 6046014
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C
; APPLICANT: Murphy, John T
; TITLE OF INVENTION: PHYTOFLUORS AS FLUORESCENT LABELS
; FILE REFERENCE: 2500.134USO UC OTT Lagarias Patent
; CURRENT APPLICATION NUMBER: US/08/904,871
; CURRENT FILING DATE: 1997-08-01
; EARLIER APPLICATION NUMBER: 60/023,217
; EARLIER FILING DATE: 1996-08-02
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1142
; TYPE: PRT
; ORGANISM: Alga (Mesocricetum)
US-08-904-871-12

Query Match 63.2%; Score 36; DB 3; Length 1142;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILH 7
|||
Db 355 AAGGILH 361

RESULT 10
US-09-346-237-5
; Sequence 5, Application US/09346237A
; Patent No. 6265197
; GENERAL INFORMATION:
; APPLICANT: Blegard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/346,237A
; CURRENT FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: PA 1998 00868
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 60/094,353
; EARLIER FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1090
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(1090)
; OTHER INFORMATION: Pullulanase
US-09-346-237-5

Query Match 61.4%; Score 35; DB 4; Length 1090;
Best Local Similarity 54.5%; Pred. No. 3e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGGILHELL 11
|||:|:|
Db 474 SASGVTHIELL 484

RESULT 11
US-08-913-014A-2
; Sequence 2, Application US/08913014A
; Patent No. 6235878
; GENERAL INFORMATION:

APPLICANT: Nishi, Kazunori
APPLICANT: Hikichi, Yukiko
APPLICANT: Shintani, Yasuohi
TITLE OF INVENTION: NOVEL FAS LIGAND-LIKE PROTEIN, ITS
TITLE OF INVENTION: PRODUCTION AND USE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: David G. Conlin, Esq.
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & USHMAN, LLP
STREET: 130 Water Street
CITY: Boston,
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913, 014A
FILING DATE: 04-SEP-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/02480
FILING DATE: July 17, 1997
ATTORNEY/AGENT INFORMATION:
NAME: David G. Conlin
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 342/47694
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 239
TYPE: Amino acid
TOPOLOGY: Linear
MOLECULE TYPE: Protein
US-08-913-014A-2
Query Match 57.9%; Score 33; DB 4; Length 239;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 3 GIGLHLE 9
DB 203 GGVVHLE 209
RESULT 12
US-08-913-014A-3
Sequence 3, Application US/08913014A
Patent No. 6235878
GENERAL INFORMATION:
APPLICANT: Nishi, Kazunori
APPLICANT: Hikichi, Yukiko
APPLICANT: Shintani, Yasuohi
TITLE OF INVENTION: NOVEL FAS LIGAND-LIKE PROTEIN, ITS
TITLE OF INVENTION: PRODUCTION AND USE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: David G. Conlin, Esq.
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & USHMAN, LLP
STREET: 130 Water Street
CITY: Boston,
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913, 014A
FILING DATE: 04-SEP-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/02480
FILING DATE: July 17, 1997
ATTORNEY/AGENT INFORMATION:
NAME: David G. Conlin
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 342/47694
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 239
TYPE: Amino acid
TOPOLOGY: Linear
MOLECULE TYPE: Protein
US-08-913-014A-3
Query Match 57.9%; Score 33; DB 4; Length 239;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 3 GIGLHLE 9
DB 203 GGVVHLE 209
RESULT 13
US-08-913-014A-1
Sequence 1, Application US/08913014A
Patent No. 6235878
GENERAL INFORMATION:
APPLICANT: Nishi, Kazunori
APPLICANT: Hikichi, Yukiko
APPLICANT: Shintani, Yasuohi
TITLE OF INVENTION: NOVEL FAS LIGAND-LIKE PROTEIN, ITS
TITLE OF INVENTION: PRODUCTION AND USE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: David G. Conlin, Esq.
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & USHMAN, LLP
STREET: 130 Water Street
CITY: Boston,
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913, 014A
FILING DATE: 04-SEP-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/02480
FILING DATE: July 17, 1997
ATTORNEY/AGENT INFORMATION:
NAME: David G. Conlin
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 342/47694
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:

INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 240
 TYPE: Amino acid
 TOPOLOGY: Linear
 MOLECULE TYPE: Protein
 US-08-913-014A-1

Query Match 57.9%; Score 33; DB 4; Length 240;
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 GGIHLLE 9
 ||:||||
 DB 204 GGIVHLE 210

RESULT 14
 US-09-072-993C-4
 Sequence 4, Application US/09072993C
 Patent No. 6346388
 GENERAL INFORMATION:
 APPLICANT: Michael R. Brigham-Burke
 APPLICANT: Peter R. Young
 TITLE OF INVENTION: A METHOD OF IDENTIFYING AGONIST AND
 FILE REFERENCE: GH-50030
 CURRENT APPLICATION NUMBER: US/09/072,993C
 CURRENT FILING DATE: 1998-05-06
 PRIOR APPLICATION NUMBER: 60/055,513
 PRIOR FILING DATE: 1997-08-13
 PRIOR APPLICATION NUMBER: 60/056,980
 PRIOR FILING DATE: 1997-08-26
 PRIOR APPLICATION NUMBER: 60/057,550
 PRIOR FILING DATE: 1997-08-29
 NUMBER OF SEQ ID NOS: 9
 SOFTWARE: FASTSEQ for Windows Version 3.0
 SEQ ID NO 4
 LENGTH: 240
 TYPE: PRT
 ORGANISM: HOMO SAPIENS
 US-09-072-993C-4

Query Match 57.9%; Score 33; DB 4; Length 240;
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 GGIHLLE 9
 ||:||||
 DB 204 GGIVHLE 210

RESULT 15
 US-08-913-014A-25
 Sequence 25, Application US/08913014A
 Patent No. 6235878
 GENERAL INFORMATION:
 APPLICANT: Mishi, Kazuo
 APPLICANT: Hikichi, Yukiko
 APPLICANT: Shtirani, Yasuaki
 TITLE OF INVENTION: NOVEL PAS LIGAND-LIKE PROTEIN, ITS
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: David G. Conlin, Esq.
 ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & USHMAN, LLP
 STREET: 130 Water Street
 CITY: Boston,
 STATE: MA
 COUNTRY: USA
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/913,014A
 FILING DATE: 04-SEP-1997
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/JP97/02480
 FILING DATE: JULY 17, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: David G. Conlin
 REGISTRATION NUMBER: 27,026
 REFERENCE/DOCKET NUMBER: 342/47694
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-523-3400
 TELEFAX: 617-523-6440
 TELEX:
 INFORMATION FOR SEQ ID NO: 25:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 242
 TYPE: Amino acid
 TOPOLOGY: Linear
 MOLECULE TYPE: Protein
 US-08-913-014A-25

Query Match 57.9%; Score 33; DB 4; Length 242;
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 GGIHLLE 9
 ||:||||
 DB 206 GGIVHLE 212

Search completed: March 20, 2003, 12:36:01
 Job time : 3.94595 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 20, 2003, 12:34:44 ; Search time 1.94595 Seconds
(without alignments)
329.684 Million cell updates/sec

Title: SEQID_1
Perfect score: 57
Sequence: 1 AAGGTLHELV 12

Scoring table:
BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 221153

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCT05_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	100.0	12	10	US-09-833-328-1
2	57	100.0	15	10	US-09-833-328-5
3	57	100.0	15	10	US-09-833-328-11
4	57	100.0	148	10	US-09-833-328-4
5	57	100.0	150	10	US-09-833-328-6
6	57	100.0	203	10	US-09-781-080B-2
7	57	100.0	242	10	US-09-978-979-1
8	57	100.0	1120	10	US-09-781-080B-11
9	57	100.0	1416	10	US-09-836-712-2
10	45	78.9	9	10	US-09-833-328-9
11	36	63.2	125	10	US-09-810-264-34
12	36	63.2	210	10	US-09-272-809-1
13	36	63.2	1096	10	US-09-833-435A-6
14	35	61.4	1090	10	US-09-833-435A-5
15	35	61.4	1091	9	US-09-262-126C-4
16	35	61.4	1091	9	US-10-245-803-4
17	34	59.6	144	9	US-10-101-464A-789
18	34	59.6	316	10	US-09-811-284-143
19	34	59.6	330	9	US-09-738-626-6014

20	34	59.6	442	9	US-10-117-417-6	Sequence 6, Appli
21	34	59.6	442	9	US-10-117-417-16	Sequence 16, Appl
22	34	59.6	547	9	US-10-101-464A-928	Sequence 928, App
23	34	59.6	1002	9	US-10-101-464A-957	Sequence 957, App
24	33	57.9	81	10	US-09-864-761-35584	Sequence 35584, A
25	33	57.9	132	9	US-09-764-868-906	Sequence 906, App
26	33	57.9	132	9	US-09-764-868-1214	Sequence 1214, Ap
27	33	57.9	159	10	US-09-731-872-402	Sequence 402, App
28	33	57.9	208	10	US-09-027-287-39	Sequence 39, Appl
29	33	57.9	208	10	US-09-252-656B-39	Sequence 39, Appl
30	33	57.9	240	10	US-09-027-287-2	Sequence 2, Appli
31	33	57.9	240	10	US-09-252-656B-2	Sequence 2, Appli
32	33	57.9	240	12	US-10-066-209-4	Sequence 4, Appli
33	33	57.9	551	13	US-09-818-143-21	Sequence 21, Appl
34	33	57.9	602	10	US-09-841-132-495	Sequence 495, App
35	33	57.9	602	10	US-09-841-132-495	Sequence 495, App
36	33	57.9	743	9	US-09-836-613-2	Sequence 565, App
37	32	56.1	38	10	US-09-864-761-36574	Sequence 36574, A
38	32	56.1	121	10	US-09-864-761-45340	Sequence 45340, A
39	32	56.1	170	9	US-10-006-856A-162	Sequence 162, App
40	32	56.1	170	10	US-09-099-882-22	Sequence 22, Appl
41	32	56.1	170	10	US-09-951-845-2	Sequence 2, Appli
42	32	56.1	284	9	US-10-028-072-242	Sequence 242, App
43	32	56.1	284	9	US-10-121-049-242	Sequence 242, App
44	32	56.1	284	9	US-10-123-904-242	Sequence 242, App
45	32	56.1	284	9	US-10-140-470-242	Sequence 242, App

ALIGNMENTS

RESULT 1
US-09-833-328-1
; Sequence 1, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Schefflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, RandoIf
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; APPLICANT: Gerritsen, Helena E.
; TITLE OF INVENTION: Compositing a Polypeptide Chain with the Amino Acid Sequence
; FILE REFERENCE: 247,00CIP
; CURRENT APPLICATION NUMBER: US/09/833,328
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 12
; TYPE: PRT
; ORGANISM: human
US-09-833-328-1
Query Match 100.0%; Score 57; DB 10; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00063;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 AAGGTLHELV 12
DB 1 AAGGTLHELV 12
RESULT 2
US-09-833-328-5
; Sequence 5, Application US/09833328

```
Patent No. US20020136713A1
GENERAL INFORMATION:
APPLICANT: Laemmle, Bernhard
APPLICANT: Schwarz, Hans-Peter
APPLICANT: Scheiflinger, Friedrich
APPLICANT: Antoine, Gerhard
APPLICANT: Kerschbaumer, RandoIf
APPLICANT: Tagliavacca, Luigina
APPLICANT: Zimmermann, Klaus
APPLICANT: Furlan, Miha
APPLICANT: Turecek, Peter
APPLICANT: Geritsen, Helena E.
TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
FILE REFERENCE: 247.00CIP
CURRENT APPLICATION NUMBER: US/09/833,328
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 09/721,254
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 15
TYPE: PRT
ORGANISM: human
US-09-833-328-5

Query Match      100.0%; Score 57; DB 10; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00079;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 AAGGILHLELLV 12
Db      1 AAGGILHLELLV 12
```

```
RESULT 3
US-09-833-328-11
Sequence 11, Application US/09833328
Patent No. US20020136713A1
GENERAL INFORMATION:
APPLICANT: Laemmle, Bernhard
APPLICANT: Schwarz, Hans-Peter
APPLICANT: Scheiflinger, Friedrich
APPLICANT: Antoine, Gerhard
APPLICANT: Kerschbaumer, RandoIf
APPLICANT: Tagliavacca, Luigina
APPLICANT: Zimmermann, Klaus
APPLICANT: Furlan, Miha
APPLICANT: Turecek, Peter
APPLICANT: Geritsen, Helena E.
TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
FILE REFERENCE: 247.00CIP
CURRENT APPLICATION NUMBER: US/09/833,328
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 09/721,254
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11
LENGTH: 15
TYPE: PRT
ORGANISM: human
US-09-833-328-11
```

```
Query Match      100.0%; Score 57; DB 10; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00079;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 AAGGILHLELLV 12
Db      1 AAGGILHLELLV 12
```

```
RESULT 4
US-09-833-328-4
Sequence 4, Application US/09833328
Patent No. US20020136713A1
GENERAL INFORMATION:
APPLICANT: Laemmle, Bernhard
APPLICANT: Schwarz, Hans-Peter
APPLICANT: Scheiflinger, Friedrich
APPLICANT: Antoine, Gerhard
APPLICANT: Kerschbaumer, RandoIf
APPLICANT: Tagliavacca, Luigina
APPLICANT: Zimmermann, Klaus
APPLICANT: Furlan, Miha
APPLICANT: Turecek, Peter
APPLICANT: Geritsen, Helena E.
TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
FILE REFERENCE: 247.00CIP
CURRENT APPLICATION NUMBER: US/09/833,328
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 09/721,254
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 148
TYPE: PRT
ORGANISM: human
US-09-833-328-4

Query Match      100.0%; Score 57; DB 10; Length 148;
Best Local Similarity 100.0%; Pred. No. 0.0081;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 AAGGILHLELLV 12
Db      1 AAGGILHLELLV 12
```

```
RESULT 5
US-09-833-328-6
Sequence 6, Application US/09833328
Patent No. US20020136713A1
GENERAL INFORMATION:
APPLICANT: Laemmle, Bernhard
APPLICANT: Schwarz, Hans-Peter
APPLICANT: Scheiflinger, Friedrich
APPLICANT: Antoine, Gerhard
APPLICANT: Kerschbaumer, RandoIf
APPLICANT: Tagliavacca, Luigina
APPLICANT: Zimmermann, Klaus
APPLICANT: Furlan, Miha
APPLICANT: Turecek, Peter
APPLICANT: Geritsen, Helena E.
TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
FILE REFERENCE: 247.00CIP
CURRENT APPLICATION NUMBER: US/09/833,328
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 09/721,254
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 150
TYPE: PRT
ORGANISM: human
US-09-833-328-6
```

```
Query Match      100.0%; Score 57; DB 10; Length 150;
Best Local Similarity 100.0%; Pred. No. 0.0082;
```


Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILHLELV 12
|||||
DB 3 AAGGILHLELV 14

RESULT 6
US-09-781-080B-2
; Sequence 2, Application US/09781080B
; Patent No. US20020142439A1
; GENERAL INFORMATION:
; APPLICANT: HOLLOWAY, JIM
; APPLICANT: SHEPPARD, PAUL
; APPLICANT: YAMAMOTO, GAYLE
; TITLE OF INVENTION: Anti-Angiogenic Intestinal peptides,
; FILE REFERENCE: 99-82
; CURRENT APPLICATION NUMBER: US/09/781,080B
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-781-080B-2

Query Match 100.0%; Score 57; DB 10; Length 203;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILHLELV 12
|||||
DB 2 AAGGILHLELV 13

RESULT 7
US-09-978-979-1
; Sequence 1, Application US/09978979
; Patent No. US20020151702A1
; GENERAL INFORMATION:
; APPLICANT: Racle, Lisa, A.
; Twine, Natalie, C.
; Agostino, Michael, J.
; Wolfman, Neil
; Morris, Elisabeth
; TITLE OF INVENTION: Aggreganase Molecules
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Home Products Corporation
; STREET: One Campus Drive
; CITY: Parsippany
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07054

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/978,979
FILING DATE: 16-Oct-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/60/241,469
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Kapinos, Ellen J.
REGISTRATION NUMBER: 32,345
REFERENCE/DOCKET NUMBER: GI 5435P
TELECOMMUNICATION INFORMATION:

TELEPHONE: (973) 660-5000
TELEFAX: (973) 683-4117
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 242 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-978-979-1

Query Match 100.0%; Score 57; DB 10; Length 242;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILHLELV 12
|||||
DB 42 AAGGILHLELV 53

RESULT 8
US-09-781-080B-11
; Sequence 11, Application US/09781080B
; Patent No. US20020142439A1
; GENERAL INFORMATION:
; APPLICANT: HOLLOWAY, JIM
; APPLICANT: SHEPPARD, PAUL
; APPLICANT: YAMAMOTO, GAYLE
; TITLE OF INVENTION: Anti-Angiogenic Intestinal Peptides,
; FILE REFERENCE: 99-82
; CURRENT APPLICATION NUMBER: US/09/781,080B
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 1120
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1120)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-781-080B-11

Query Match 100.0%; Score 57; DB 10; Length 1120;
Best Local Similarity 100.0%; Pred. No. 0.063;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGGILHLELV 12
|||||
DB 105 AAGGILHLELV 116

RESULT 9
US-09-836-712-2
; Sequence 2, Application US/09836712
; Patent No. US20010049106A1
; GENERAL INFORMATION:
; APPLICANT: PFIZER INC.
; TITLE OF INVENTION: ADAMTS POLYPEPTIDES, NUCLEIC ACIDS ENCODING THEM, AND
; FILE REFERENCE: PC10851A
; CURRENT APPLICATION NUMBER: US/09/836,712
; CURRENT FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1416
; TYPE: PRT
; ORGANISM: Human
US-09-836-712-2

Query Match 100.0%; Score 57; DB 10; Length 1416;
Best Local Similarity 100.0%; Pred. No. 0.08;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLELLV 12
| | | | | | | | | | | | | |
Db 98 AAGGILHLELLV 109

RESULT 10
US-09-833-328-9
; Sequence 9, Application US/09833328
; Patent No. US20020136713A1
; GENERAL INFORMATION:
; APPLICANT: Laemmle, Bernhard
; APPLICANT: Schwarz, Hans-Peter
; APPLICANT: Scheiflinger, Friedrich
; APPLICANT: Antoine, Gerhard
; APPLICANT: Kerschbaumer, Rndolf
; APPLICANT: Tagliavacca, Luigina
; APPLICANT: Zimmermann, Klaus
; APPLICANT: Furlan, Miha
; APPLICANT: Turecek, Peter
; TITLE OF INVENTION: Composition Exhibiting a von Willebrand Factor (vWF) Protease Act
; TITLE OF INVENTION: Comprising a Polypeptide Chain with the Amino Acid Sequence AAGG
; FILE REFERENCE: 247.00CIP
; CURRENT APPLICATION NUMBER: US/09/833,328
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/721,254
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 9
; TYPE: PRT
; ORGANISM: human
US-09-833-328-9

Query Match 78.9%; Score 45; DB 10; Length 9;
Best Local Similarity 100.0%; Pred. No. 2e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILHLE 9
| | | | | | | | | |
Db 1 AAGGILHLE 9

RESULT 11
US-09-810-264-34
; Sequence 34, Application US/09810264
; Patent No. US20020076775A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Virginia C.
; APPLICANT: Ramodu, Omolayo O.
; APPLICANT: Hu, Xu
; APPLICANT: Zhang, Lingyu
; TITLE OF INVENTION: WRKY Transcription Factors and Methods
; TITLE OF INVENTION: of Use
; FILE REFERENCE: 1183
; CURRENT APPLICATION NUMBER: US/09/810,264
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/190,467
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Zea mays
US-09-810-264-34

Query Match 63.2%; Score 36; DB 10; Length 125;
Best Local Similarity 66.7%; Pred. No. 27;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGILHLELLV 12
| | | | | | | | | | | | | |
Db 24 AAGGILHLELLV 35

RESULT 12
US-09-272-809-11
; Sequence 11, Application US/09272809
; Patent No. US20020022239A1
; GENERAL INFORMATION:
; APPLICANT: Lagarias, John C.
; TITLE OF INVENTION: Phytocfluors as fluorescent labels
; FILE REFERENCE: 2500.118US0
; CURRENT APPLICATION NUMBER: US/09/272,809
; CURRENT FILING DATE: 1998-03-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Mcphylb
US-09-272-809-11

Query Match 63.2%; Score 36; DB 10; Length 210;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGGILH 7
| | | | | | | |
Db 153 AAGGILH 159

RESULT 13
US-09-833-435A-6
; Sequence 6, Application US/09833435A
; Patent No. US20020081670A1
; GENERAL INFORMATION:
; APPLICANT: Bisgard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629.200-US
; CURRENT APPLICATION NUMBER: US/09/833,435A
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/346,237
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: 60/094,353
; PRIOR FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1096
; TYPE: PRT
; ORGANISM: Klebsiella aerogenes
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(1096)
; OTHER INFORMATION: Pullulanase
US-09-833-435A-6

Query Match 63.2%; Score 36; DB 10; Length 1096;
Best Local Similarity 54.5%; Pred. No. 2.5e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGILHLELL 11
| | | | | | | | | |
Db 484 SASGVTHLELL 494

```

RESULT 14
US-09-833-435A-5
; Sequence 5, Application US/09833435A
; Patent No. US20020081670A1
; GENERAL INFORMATION:
; APPLICANT: Bisgaard-Frantzen, Henrik
; APPLICANT: Svendsen, Allan
; TITLE OF INVENTION: Starch Debranching Enzymes
; FILE REFERENCE: 5629, 200-US
; CURRENT APPLICATION NUMBER: US/09/833,435A
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 09/346,237
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: 60/094,353
; PRIOR FILING DATE: 1998-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 1090
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(1090)
; OTHER INFORMATION: pullulanase
US-09-833-435A-5

```

```

Query Match      61.4%; Score 35; DB 10; Length 1090;
Best Local Similarity 54.5%; Pred. No. 3.6e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAGGILHLELL 11
   |||:|:|
Db 474 SASGVTHVELL 484

```

```

RESULT 15
US-09-262-126C-4
; Sequence 4, Application US/09262126C
; Publication No. US20030013180A1
; GENERAL INFORMATION:
; APPLICANT: Miller, Brian S.
; APPLICANT: Shetty, Jayarama K.
; TITLE OF INVENTION: Modified Forms of Pullulanase
; FILE REFERENCE: GC396-2
; CURRENT APPLICATION NUMBER: US/09/262,126C
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1091
; TYPE: PRT
; ORGANISM: Klebsiella pneumonia
US-09-262-126C-4

```

```

Query Match      61.4%; Score 35; DB 9; Length 1091;
Best Local Similarity 54.5%; Pred. No. 3.6e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAGGILHLELL 11
   |||:|:|
Db 475 SASGVTHVELL 485

```

Search completed: March 20, 2003, 12:36:34
Job time : 3.94595 secs